

**Item und Skala:**

**Empirische Untersuchungen zur Gültigkeit psychologischer Messungen  
anhand physikalischer Merkmale**

**Inauguraldissertation**

zur Erlangung des Doktorgrades der Philosophie an der  
Fakultät für Psychologie und Pädagogik der  
Ludwig-Maximilians-Universität München



**vorgelegt von**

**Kathryn Eichhorn**

**aus Kelheim**

**2019**

Betreuer der Dissertation:

Erstgutachter:

Prof. Dr. Markus Bühner

Zweitgutachter:

Prof. Dr. Sven Hilbert

Datum der mündlichen Prüfung:

07. Februar 2019

*„Die menschliche Sprache gleicht einem zersprungenen Kessel,  
auf den wir krude Rhythmen wie für Tanzbären trommeln,  
während wir uns danach sehnen,  
eine Musik zu machen,  
bei der die Sterne schmelzen.“*

— Gustave Flaubert, *Madame Bovary*

In dieser Arbeit wird sich sprachlich auf die männliche Form personenbezogener Substantive beschränkt. Dies impliziert keine Geschlechtsdiskriminierung oder Verletzung des Gleichheitsgrundsatzes, sondern dient ausschließlich der sprachlichen Vereinfachung zur besseren Lesbarkeit.

## **Zusammenfassung**

Die Bestimmung der Validität psychologischer Fragebögen stellt vor allem deshalb eine Herausforderung dar, weil die wahren Werte latenter psychologischer Konstrukte immer unbekannt sind. Die Gültigkeit eines Fragebogens muss mit Behelfslösungen und Korrelationen bestimmt werden. Diese Validierungsansätze können, beispielsweise wegen Zirkelschlüssen oder Post-hoc-Erklärungen, problematisch sein. Aus diesem Grund untersuchte die vorliegende Arbeit anhand physikalischer Merkmale verschiedene Aspekte der Gültigkeit psychologischer Fragebögen. Physikalische Merkmale wurden mittels Items erfasst, wodurch den psychologischen Messungen reale Werte physikalischer Messgrößen gegenübergestellt werden konnten.

Die Studien I und II untersuchten die psychometrische Güte einzelner Items, die physikalisch messbare Personenmerkmale (Alter, Körpergröße, Körpergewicht und Einkommen) erfassen. Zentrale Fragestellung war, ob Bildungsniveau und relatives Körpergewicht (Studie I) sowie Itemformulierung, Skalenformat und Selbstbewertungsperspektive (Studie II) Einflussvariablen sind, die dazu führen, dass die Relationen der realen Werte im Mittel unterschiedlich gut auf der Antwortskala erhalten bleiben. Es zeigte sich, dass die Selbsteinschätzung unabhängig vom Bildungsniveau blieb (Studien I und II). Bei nach Geschlecht getrennter Datenauswertung lieferte das relative Körpergewicht keinen inkrementellen Beitrag für die Vorhersage der Selbsteinschätzung des Körpergewichts über das absolute Körpergewicht hinaus (Studie I). Itemformulierung, Skalenverankerung und Skalenbreite hatten keinen eindeutigen Einfluss auf die Selbstbewertung (Studie II). Die Selbstbewertungsperspektive kam vor allem bei stark von subjektiven Faktoren bestimmten Merkmalen zum Tragen (Studie II).

In Studie III wurde der Zusammenhang zwischen den Merkmalen Alter, Körpergröße und Körpergewicht und deren Abbildung auf psychologische Skalen (sechsstufiges oder dichotomes Antwortformat) untersucht. Im Mittelpunkt der Analysen stand die Frage, wie gut die Skalen in der Lage sind, die physikalischen Merkmale abzubilden. Beim Merkmal Körpergröße ergaben sich die höchsten Reliabilitätsschätzungen und die größten aufgeklärten Varianzanteile. Es zeigte sich kein eindeutiges Muster, ob durch Summenwerte, Faktorwerte oder Personenparameter mehr Varianz aufgeklärt werden kann

und ob mit einer dichotomen oder polytomen Antwortskala die realen Werte in den Regressionsanalysen besser vorhergesagt werden können.

Zusammenfassend legen die drei Studien nahe, dass Messungen physikalischer Merkmale mit psychologischen Items bzw. Skalen grundsätzlich möglich sind, deren Validität jedoch vor allem bei stark von subjektiven Faktoren beeinflussten Merkmalen deutlich abnimmt.

## **Inhaltsverzeichnis**

Zusammenfassung .....	iv
Abbildungsverzeichnis .....	ix
Tabellenverzeichnis.....	xi
Abkürzungsverzeichnis .....	xiv
1. Allgemeine Einführung.....	15
2. Studie I: Einfluss von Bildungsniveau und relativem Körpergewicht auf das Skalenniveau23	
2.1 Theorie Studie I .....	23
2.1.1 Einführung .....	23
2.1.2 Der Referenzrahmen als Einflussgröße .....	24
2.1.3 Die graphische Darstellung statistischer Informationen.....	25
2.1.4 Rationale.....	26
2.2 Methoden Studie I .....	28
2.2.1 Stichprobe .....	28
2.2.2 Material .....	29
2.2.3 Erhebung .....	31
2.2.4 Analysen .....	32
2.3 Ergebnisse Studie I .....	34
2.3.1 Deskriptive Statistik.....	34
2.3.2 Regressionsanalyse .....	37
2.4 Diskussion Studie I.....	51
2.4.1 Zusammenfassung der Ergebnisse .....	51
2.4.2 Interpretation und Implikationen der Ergebnisse.....	51
2.4.3 Limitationen und Ausblick.....	55
2.4.4 Fazit .....	55
3. Studie II: Einfluss von Itemformulierung und Skalenformat auf das Skalenniveau.....	57
3.1 Theorie Studie II .....	57
3.1.1 Einführung .....	57
3.1.2 Negativ gepolte und negativ formulierte Items.....	58
3.1.3 Antwortskalen mit festen Intervallen .....	61
3.1.4 Rationale.....	64
3.2 Methoden Studie II .....	66

3.2.1 Präregistrierung.....	66
3.2.2 Stichprobe .....	66
3.2.3 Material .....	67
3.2.4 Erhebung .....	68
3.2.5 Analysen .....	69
3.3 Ergebnisse Studie II .....	73
3.3.1 Deskriptive Statistik.....	73
3.3.2 Regressionsanalyse .....	79
3.3.3 Ergebnisse im Kontext der Forschungsfragen.....	102
3.4 Diskussion Studie II.....	102
3.4.1 Zusammenfassung der Ergebnisse .....	102
3.4.2 Interpretation und Implikationen der Ergebnisse.....	103
3.4.3 Limitationen und Ausblick .....	109
3.4.4 Fazit .....	110
4. Studie III: Konstruktion einer psychometrischen Skala für physikalische Merkmale.....	112
4.1 Theorie Studie III .....	112
4.1.1 Einführung .....	112
4.1.2 Fragebogen als Messverfahren .....	113
4.1.3 Testtheoretische Überlegungen.....	115
4.1.4 Rationale.....	117
4.2 Methoden Studie III .....	118
4.2.1 Präregistrierung.....	118
4.2.2 Stichprobe .....	119
4.2.3 Skalenkonstruktion.....	120
4.2.4 Material .....	125
4.2.5 Erhebung .....	125
4.2.6 Analysen .....	126
4.3 Ergebnisse Studie III .....	130
4.3.1 Deskriptive Statistik.....	130
4.3.2 Korrelations-, Reliabilitäts- und Itemanalyse .....	133
4.3.3 Explorative und Konfirmatorische Faktorenanalyse .....	137
4.3.4 Probabilistische Testmodelle .....	140

4.3.5 Regressionsanalyse .....	143
4.3.6 Ergebnisse im Kontext der Forschungsfragen.....	144
4.4 Diskussion Studie III.....	146
4.4.1 Zusammenfassung der Ergebnisse .....	146
4.4.2 Interpretation und Implikationen der Ergebnisse.....	147
4.4.3 Limitationen und Ausblick .....	153
4.4.4 Fazit .....	155
5. Abschließende Diskussion.....	156
Literaturverzeichnis .....	164
Anhang .....	191
A. Anhang zu Studie I .....	191
B. Anhang zu Studie II.....	206
C. Anhang zu Studie III.....	218



## Abbildungsverzeichnis

Abbildung 2.3.1. Regression Antwortskala und reale Werte für das Merkmal Körpergewicht für Frauen .....	47
Abbildung 2.3.2. Regression Antwortskala und reale Werte für das Merkmal Körpergewicht für Männer .....	48
Abbildung 3.3.1. Regression Antwortskala und reale Werte für das Merkmal Alter für Frauen .....	83
Abbildung 3.3.2. Regression Antwortskala und reale Werte für das Merkmal Alter für Männer.....	84
Abbildung 3.3.3. Regression Antwortskala und reale Werte für das Merkmal Einkommen für Frauen.....	85
Abbildung 3.3.4. Regression Antwortskala und reale Werte für das Merkmal Einkommen für Männer .....	86
Abbildung 3.3.5. Regression Antwortskala und reale Werte für das Merkmal Körpergröße für Frauen.....	87
Abbildung 3.3.6. Regression Antwortskala und reale Werte für das Merkmal Körpergröße für Männer .....	88
Abbildung 3.3.7. Regression Antwortskala und reale Werte für das Merkmal Körpergewicht für Frauen.....	89
Abbildung 3.3.8. Regression Antwortskala und reale Werte für das Merkmal Körpergewicht für Männer .....	90
Abbildung 3.3.9. Regression Antwortskala und reale Werte für die Merkmale Alter, Einkommen, Körpergröße und Körpergewicht für Frauen – Egozentrische Perspektive.....	98
Abbildung 3.3.10. Regression Antwortskala und reale Werte für die Merkmale Alter, Einkommen, Körpergröße und Körpergewicht für Frauen – Allozentrische Perspektive .....	99
Abbildung 3.3.11. Regression Antwortskala und reale Werte für die Merkmale Alter, Einkommen, Körpergröße und Körpergewicht für Männer – Egozentrische Perspektive ....	100
Abbildung 3.3.12. Regression Antwortskala und reale Werte für die Merkmale Alter, Einkommen, Körpergröße und Körpergewicht für Männer – Allozentrische Perspektive....	101

Abbildung 4.3.1. Item Characteristic Curves für Item 6 („Ich mache mir viele Gedanken über das Älterwerden.“) als Beispiel für ungeordnete Schwellen.....	142
Abbildung 4.3.2. Item Characteristic Curves für Item 27 („Bei Hosen habe ich eine große Bundweite.“) als Beispiel für geordnete Schwellen.....	142

## Tabellenverzeichnis

Tabelle 2.2.1. Verteilung von Geschlecht, Bildungsniveau und Alter innerhalb der Bedingungen GV, GSV, GV/GSV und KV.....	29
Tabelle 2.3.1. Mittelwert, Standardabweichung und Median der realen Werte getrennt nach Bildungsniveau und Geschlecht – Bedingung GV .....	35
Tabelle 2.3.2. Mittelwert, Standardabweichung und Median der realen Werte getrennt nach Bildungsniveau und Geschlecht – Bedingung GSV.....	35
Tabelle 2.3.3. Mittelwert, Standardabweichung und Median der realen Werte getrennt nach Bildungsniveau und Geschlecht – Bedingung GV/GSV .....	36
Tabelle 2.3.4. Mittelwert, Standardabweichung und Median der realen Werte getrennt nach Bildungsniveau und Geschlecht – Bedingung KV .....	36
Tabelle 2.3.5. Ergebnisse Regressionsanalysen mit Bildungsniveau für Frauen und Männer – Bedingung GV .....	38
Tabelle 2.3.6. Ergebnisse Regressionsanalysen mit Bildungsniveau für Frauen und Männer – Bedingung GSV .....	40
Tabelle 2.3.7. Ergebnisse Regressionsanalysen mit Bildungsniveau für Frauen und Männer – Bedingung GV/GSV.....	42
Tabelle 2.3.8. Ergebnisse Regressionsanalysen mit Bildungsniveau für Frauen und Männer – Bedingung KV .....	44
Tabelle 2.3.9. Schwellenparameter und Steigung für Körpergewicht der Bedingung GV für Frauen und Männer.....	49
Tabelle 2.3.10. Schwellenparameter und Steigung für Körpergewicht der Bedingung GSV für Frauen und Männer.....	49
Tabelle 2.3.11. Schwellenparameter und Steigung für Körpergewicht der Bedingung GV/GSV für Frauen und Männer.....	50
Tabelle 2.3.12. Schwellenparameter und Steigung für Körpergewicht der Bedingung KV für Frauen und Männer.....	50
Tabelle 3.2.1. Verteilung von Geschlecht, Bildungsniveau und Alter innerhalb der Bedingungen 5PS, 4PS und 2PS .....	67

Tabelle 3.3.1. Mittelwert, Standardabweichung, Median, Minimum und Maximum reale Werte getrennt für Frauen und Männer .....	73
Tabelle 3.3.2. Mittelwert, Standardabweichung und Median der realen Werte getrennt nach Bildungsniveau und Geschlecht – Bedingung 5PS .....	75
Tabelle 3.3.3. Mittelwert, Standardabweichung und Median der realen Werte getrennt nach Bildungsniveau und Geschlecht – Bedingung 4PS .....	75
Tabelle 3.3.4. Mittelwert, Standardabweichung und Median der realen Werte getrennt nach Bildungsniveau und Geschlecht – Bedingung 2PS .....	76
Tabelle 3.3.5. Mittelwert, Standardabweichung und Median beobachtete Werte für Frauen.....	77
Tabelle 3.3.6. Mittelwert, Standardabweichung und Median beobachtete Werte für Männer .....	78
Tabelle 3.3.7. Ergebnisse Spearman- bzw. punktbiseriale Korrelationen der Bedingungen 5PS, 4PS und 2PS für Frauen und Männer .....	79
Tabelle 3.3.8. Schwellenparameter und Steigung der Bedingung 5PS für Frauen und Männer .....	91
Tabelle 3.3.9. Schwellenparameter und Steigung der Bedingung 4PS für Frauen und Männer .....	92
Tabelle 3.3.10. Ergebnisse logistische Regression der Bedingung 2PS für Frauen und Männer .....	93
Tabelle 3.3.11. Ergebnisse Regressionsanalysen mit Bildungsniveau für Frauen und Männer – Bedingung 5PS .....	94
Tabelle 3.3.12. Ergebnisse Regressionsanalysen mit Bildungsniveau für Frauen und Männer – Bedingung 4PS .....	95
Tabelle 3.3.13. Ergebnisse Regressionsanalysen mit Bildungsniveau für Frauen und Männer – Bedingung 2PS .....	96
Tabelle 4.2.1. Verteilung von Geschlecht, Bildungsniveau, Muttersprache und Alter innerhalb der Bedingungen 2PS und 6PS .....	120
Tabelle 4.3.1. Mittelwert, Standardabweichung, Median, Minimum und Maximum reale Werte getrennt für Frauen und Männer – Bedingung 2PS .....	131

Tabelle 4.3.2. Mittelwert, Standardabweichung, Median, Minimum und Maximum reale Werte getrennt für Frauen und Männer – Bedingung 6PS .....	131
Tabelle 4.3.3. Mittelwert, Standardabweichung und Median der realen Werte getrennt nach Bildungsniveau und Geschlecht – Bedingung 2PS .....	132
Tabelle 4.3.4. Mittelwert, Standardabweichung und Median der realen Werte getrennt nach Bildungsniveau und Geschlecht – Bedingung 6PS .....	133
Tabelle 4.3.5. Reliabilitäten für die Bedingungen 2PS und 6PS .....	135
Tabelle 4.3.6. Mittelwerte, Standardabweichungen und Trennschärfen für die Bedingungen 2PS und 6PS .....	136
Tabelle 4.3.7. Explorative Faktorenanalyse mit drei Faktoren (Maximum Likelihood, Promax) – Bedingungen 2PS und 6PS .....	138
Tabelle 4.3.8. Faktorkorrelationen EFA mit drei Faktoren (Maximum Likelihood, Promax) – Bedingungen 2PS und 6PS .....	139
Tabelle 4.3.9. Ergebnisse einfache lineare Regressionen für Alter, Körpergröße und Körpergewicht – Bedingungen 2PS und 6PS .....	144

## Abkürzungsverzeichnis

1PL-Modell	Einparametrisches logistisches Modell
2PL-Modell	Zweiparametrisches logistisches Modell
2PS	Dichotome Antwortskala
4PS	Vierstufige Antwortskala
5PS	Fünfstufige Antwortskala
6PS	Sechsstufige Antwortskala
AIC	Akaike Information Criterion
BIC	Bayes Information Criterion
BMI	Body Mass Index
CFA	Konfirmatorische Faktorenanalyse
CFI	Comparative fit index
EFA	Explorative Faktorenanalyse
GPCM	Generalisiertes-Partial-Credit-Modell
GSV	Anzeige Balkendiagramm mit der geschlechtsspezifischen Merkmalsverteilung in der Population
GV	Anzeige Balkendiagramm mit der Gesamtmerkmalsverteilung in der Population
GV/GSV	Gleichzeitige Anzeige Balkendiagramme mit der Gesamtmerkmalsverteilung und der geschlechtsspezifischen Merkmalsverteilung in der Population
kEUR	Tausend Euro
KTT	Klassische Testtheorie
KV	Anzeige keiner Merkmalsverteilung
LMU	Ludwig-Maximilians-Universität München
NEO-FFI	NEO-Fünf-Faktoren-Inventar
NEO-PI-R	NEO-Persönlichkeitsinventar, revidierte Fassung
PCM	Partial-Credit-Modell
PTT	Probabilistische Testtheorie
RMSEA	Root mean square error of approximation
SRMR	Standardized root mean residual
WLSMV	Weighted least squares means and variance adjusted

## 1. Allgemeine Einführung

Lassen sich psychische Phänomene *messen*? Schon in frühester Menschheitsgeschichte gab es die Überzeugung, dass sich Menschen in Persönlichkeit, kognitiven Fähigkeiten und Verhalten unterscheiden und dass diese Unterschiede erfasst werden können. So wurden in der westlichen Antike interindividuelle Unterschiede, eng verbunden mit der Physiologie, in der antiken Typenlehre erfasst und beschrieben. Hippokrates (460 – 377 v. Chr.) vertrat in seiner Vier-Säfte-Lehre die Auffassung, dass Menschen aufgrund unterschiedlicher Mischungen von Körpersäften differieren (Deister, 2009). Aristoteles (384 – 322 v. Chr.) glaubte, in der Blutbeschaffenheit verschiedene Temperamentstypen wiederzufinden. Platons (427 – 347 v. Chr.) psychologische Auffassung bildete dazu ein Kontrastprogramm. Er erkannte zwar interindividuelle Unterschiede an, jedoch hatte für ihn die Seele einen göttlichen Ursprung, weshalb es schwierig sei, sie streng zu definieren und zu erfassen (Aiken & Groth-Marnat, 2006; Eckardt, 2010).

Die alt-chinesische Kultur war den Denkern der westlichen Antike weit voraus. Bereits 2 200 v. Chr. existierten in China rudimentäre psychologische Testformen, mit denen chinesische Kaiser die Eignung ihrer Beamten überprüfen ließen. Dabei handelte es sich zunächst eher um mündliche als um schriftliche Prüfungen, bei denen nicht nur die Antworten der Prüflinge beurteilt wurden, sondern auch, wie sie diese gegeben hatten (Bowman, 1989).

Im europäischen Mittelalter bestand zumindest aus wissenschaftlicher Sicht kaum Interesse an interindividuellen Unterschieden und deren Erfassung. Das Leben der Menschen wurde weitestgehend durch die soziale Schicht bestimmt, in die sie geboren wurden. Für individuelle Entwicklung und Ausdruck gab es wenig Freiheiten. Erst in der Renaissance wurde die bereits seit der Antike bekannte Idee wiederbelebt, dass Menschen einzigartig sind. Dennoch entstanden erst gegen Ende des 19. Jahrhunderts psychologische Testverfahren zur wissenschaftlichen Erforschung individueller Unterschiede in Fähigkeiten und Persönlichkeit (Aiken & Groth-Marnat, 2006).

Zuvor war jedoch noch eine andere Debatte notwendig: Seit dem Zeitalter der Aufklärung wurde die Frage der Wissenschaftlichkeit der Psychologie diskutiert, bei der die Messbarkeit psychischer Phänomene und damit deren Mathematisierbarkeit den zentralen Aspekt darstellte (Eckardt, 2010). Vorherrschende Meinung war, dass Ziel jeglicher

Psychologie das Verständnis individueller Personen sei. Dieses Verständnis beruhe auf in irgendeiner Art und Weise zustande gekommenen Erfahrungen. Große Denker ihrer Zeit, wie Pascal (1623 – 1662) oder Kant (1724 – 1804), waren sich darin einig, dass das psychische Erleben zu komplex sei, um es mit mathematisch-formalen Gesetzmäßigkeiten beschreiben zu können. Kant vertrat die Annahme, dass die Psychologie nicht mathematisierbar sei und genau darin die Unmöglichkeit lag, Psychologie als exakte Wissenschaft zu betreiben (Eckardt, 2010; Klemm, 1911). Vielleicht unterschätzten Pascal und Kant die Möglichkeiten der Mathematik, da beispielsweise wichtige Aspekte der Wahrscheinlichkeitstheorie zu ihrer Zeit noch nicht entwickelt waren. So führte Kolmogorow (1903 – 1987) erst 1933 in dem Werk *Grundbegriffe der Wahrscheinlichkeitsrechnung* die Axiome der Wahrscheinlichkeitstheorie ein, welche das Problem aus der Welt schafften, dass die Wahrscheinlichkeit über die Wahrscheinlichkeit definiert war (Brückler, 2018).

Als „Pionier der mathematischen Psychologie“ (Eckardt, 2010, S. 50) wird häufig Herbart (1776 – 1841) angeführt. Herbart nahm an, dass psychische Phänomene bestimmten physikalischen Gesetzen unterliegen, die er in komplizierten mathematischen Formeln formulierte. Jedoch basierten diese Formeln nicht auf *Messungen*, sondern auf eigenen Mutmaßungen (Eckardt, 2010). Deshalb wird der Herbartsche Mathematisierungsversuch eher als „willkürlich axiomatische Annahme“ (Windelband & Heimsoeth, 1957; zitiert nach Eckardt, 2010, S. 50) angesehen, dessen wissenschaftsgeschichtlicher Wert fragwürdig ist (Pongratz, 1984).

Fechner (1801 – 1887) wagte den Schritt in Richtung Wissenschaftlichkeit. Ihm ging es ausdrücklich um ein *psychisches Maß*. Ihm zufolge müssten in der Psychologie zur Physiologie vergleichbare Standards etabliert werden, um eine wissenschaftliche Betrachtung psychologischer Probleme zu ermöglichen. Dafür wäre die Erfüllung von zwei grundlegenden Voraussetzungen notwendig: Die Einführung von Maß und Zahl sowie die Anwendung des Experiments (Eckardt, 2010). Fechners Überzeugung, dass psychische Phänomene grundsätzlich messbar sind bzw. der Messbarkeit zugänglich gemacht werden können, stützte sich sowohl auf forschungsmethodische Anregungen als auch auf naturphilosophische Axiome. Im Wesentlichen sind dies die sinnesphysiologischen Schwellenuntersuchungen Webers, der Energiesatz von Helmholtz und das Axiom des psychophysischen Parallelismus bei Leibniz (für einen Überblick vgl. Gundlach, 1993).



Wundt (1832 – 1920) betonte in seinen *Beiträgen zur Theorie der Sinneswahrnehmung*, dass die Entwicklung einer gegenstandsspezifischen Methodik essentiell für die Begründung einer wissenschaftlichen Psychologie sei. Als dafür notwendiges Hilfsmittel nannte er u. a. die Statistik (Graumann, 1980). Heute gilt Wundt als Begründer der eigenständig-wissenschaftlichen Psychologie (Schmithüsen & Krampen, 2015).

Im Jahr 1932 wurde ein Ausschuss der *Britischen Vereinigung zur Förderung der Wissenschaft* berufen, um sich erneut der eingangs gestellten Frage zu widmen. So sollten Wissenschaftler der Disziplinen Mathematik und Psychologie die Möglichkeit quantitativer Messungen menschlichen Erlebens prüfen und darüber berichten. In einem Zwischenbericht im Jahr 1938 beschwerte sich ein Teilnehmer über seine Kollegen insofern „that his colleagues came out by that same door as they went in“ (zitiert nach Narens, 2007, S. 61). Nach sieben Jahren wurde das Komitee ohne Einigung aufgelöst (Narens, 2007).

Stevens (1946) begründete das Scheitern des Komitees damit, dass kein Konsens über die Bedeutung des Messens erzielt werden konnte, woraufhin er seinen einflussreichen Artikel zu Skalenniveaus veröffentlichte. Damit führte er eine Messtheorie in die Psychologie ein, die von Krantz, Luce, Suppes und Tversky (1971) zur Blüte gebracht wurde (Saint-Mont, 2011). Stevens (1946) definierte vier verschiedene Niveaus von Messskalen mit zugehörigen gültigen statistischen Analysen sowie deren Implikationen für psychologische Messungen: Nominal-, Ordinal-, Intervall- und Verhältnisskala. Er löste damit anhaltende Kontroversen aus, auf welchem Skalenniveau welche statistischen Analysen Gültigkeit besitzen (z. B. Gardner, 1975; Knapp, 1990). Dabei folgen einige Autoren seiner Annahme, dass parametrische Analysen, beispielsweise die Berechnungen von Standardabweichung und Mittelwert, erst ab Intervallskalenniveau vorgenommen werden dürfen. Eine Verletzung dieser Annahme könne zu inkorrekten Ergebnissen und dadurch zu falschen Interpretationen und Schlussfolgerungen führen (z. B. Kuzon, Urbanek & McCabe, 1996; Siegel & Castellan, 1988; Stevens, 1946). Gegenstimmen argumentieren, dass die Anwendung parametrischer Methoden bei nicht-metrisch skalierten Daten die Ergebnisse nicht oder nur marginal verzerrten. Das tatsächlich vorliegende Messniveau könne vernachlässigt werden, da die meisten statistischen Methoden robust gegenüber Verletzungen der Annahme des Intervallskalenniveaus seien. Der praktische Nutzen dieser theoretischen Annahmen sei zweifelhaft, da man auch bei deren Verletzung oft trotzdem

empirisch sinnvolle Ergebnisse erhalten (z. B. Baker, Hardyck & Petrinovich, 1966; Gaito, 1960; Labovitz, 1967; Lord, 1953; Norman, 2010). Die Frage, ob parametrische Analysen für nicht-metrisch skalierte Daten angemessen oder unangemessen sind, führte zu einer Polarisierung der Meinungen mit zeitweise heftigen Debatten (z. B. Hand, 1996; Velleman & Wilkinson, 1993). Dies ist nicht überraschend, da mathematisch hergeleitete Sätze samt ihren normativen Implikationen auf empirische Fakten treffen. Letztendlich geht es um die Entscheidung zwischen der Verteidigung messtheoretischer Grundlagen und der Empirie (Saint-Mont, 2011).

Die heutige akademische Psychologie versteht unter Messen die homomorphe, d. h. strukturgetreue Abbildung eines empirischen Relativs auf ein numerisches Relativ. Die zwischen den Objekten bestehenden empirischen Beziehungen müssen sich in entsprechend zugewiesenen Nummern eindeutig widerspiegeln. Jede empirische Größe muss exakt einer numerischen Größe zugeordnet werden (Bühner & Ziegler, 2017; Krantz et al., 1971). Der Übergang vom empirischen Relativ in das numerische Relativ ist die Skala, mit deren Hilfe Messungen vorgenommen werden. Sie ist eine Zuordnungsvorschrift von Zahlenwerten zu Messobjekten (Bühner, 2011).

Verschiedene empirisch existierende Relationen wie Unterschiedlichkeit, Rangfolge, Intervalle und Verhältnisse können bei der Messung von Eigenschaften erhalten bleiben (Bortz & Schuster, 2010). Daraus ergeben sich verschiedene Skalenniveaus, wobei die von Stevens (1946) beschriebenen Skalenniveaus die vier wichtigsten sind (Nominal-, Ordinal-, Intervall- und Verhältnisskala). Bei der Nominalskala ist lediglich eine Aussage zu Unterschiedlichkeit bzw. Gleichheit möglich, bei der Ordinalskala außerdem zu Größer-Kleiner-Relationen, bei der Intervallskala des Weiteren zur Gleichheit von Differenzen und bei der Verhältnisskala zudem zur Gleichheit von Verhältnissen (Bortz & Schuster, 2010; Stevens, 1946).

Viele statistische Tests setzen voraus, dass die untersuchten Variablen mindestens Intervallskalenniveau aufweisen. Die Abbildungsvorschrift muss also sicherstellen, dass Personen mit denselben Merkmalsausprägungen dieselben Zahlen erhalten, höheren Merkmalsausprägungen höhere Zahlen zugeordnet werden und Messwertdifferenzen sinnvoll interpretierbar sind (Bühner & Ziegler, 2017). Demzufolge müsste eine gelungene

Messung diese Relationen auf der verwendeten Ratingskala abbilden und diese ebenfalls Intervallskalenniveau aufweisen (Bühner, 2011).

Fragebögen bzw. Skalen sind die wichtigsten und meistangewendeten psychologischen Test- bzw. Messverfahren (Kallus, 2016; Mummendey & Grau, 2014). Bei ihrer Konstruktion muss gewährleistet sein, dass Merkmalsrelationen richtig abgebildet werden und das erfasst wird, was erfasst werden soll (Bühner, 2011; Eid & Schmidt, 2014). Die Anforderungen sind dabei ähnlich wie bei Merkmalsmessungen in anderen wissenschaftlichen Disziplinen. So muss jede wissenschaftliche Messmethode allgemein anerkannte und bewährte Gütekriterien erfüllen (Kersting, 2008; Kersting, Häcker & Hornke, 2011). Die drei wichtigsten Anforderungen sind *Objektivität*, d. h. ein Verfahren muss unabhängig von Durchführung, Auswertung und Interpretation sein, *Reliabilität*, d. h. ein Verfahren sollte möglichst wenig messfehlerbehaftet sein und *Validität*, d. h. ein Verfahren sollte das tatsächlich gewünschte Merkmal erfassen (Bühner, 2011). Eid und Schmidt (2014) veranschaulichen die Gemeinsamkeit psychologischer und physikalischer Messungen am Beispiel einer Längenmessung: Zur Bestimmung der Körpergröße einer erwachsenen Person ist ein Messapparat notwendig, der diese in einer adäquaten Maßeinheit, beispielsweise Zentimetern, misst und nicht eine andere Eigenschaft erfasst, wie beispielsweise Körpergewicht (Validität). Dieser Messapparat soll die Eigenschaft haben, dass wiederholte Messungen am selben Erwachsenen zu (idealerweise) identischen Ergebnissen führen (Reliabilität). Der Messvorgang sollte so gestaltet sein, dass er durch die Person, die die Messung durchführt, nicht systematisch verändert werden kann (Objektivität).

Psychologische Messungen verfolgen im Grunde dasselbe Ziel und folgen ähnlichen Regeln wie physikalische Messungen (Eid & Schmidt, 2014). Dennoch unterscheiden sie sich in einem grundlegenden Aspekt. Sind Reliabilität und Objektivität an den Messvorgang selbst geknüpft, wird zur Klärung der Validität üblicherweise ein Vergleichsstandard herangezogen. Bei physikalischen Messungen geschieht dies beispielsweise im Rahmen der Eichung eines Messinstruments, bei der fest definierte und allgemein akzeptierte Standards verwendet werden. So ist die international übliche Basiseinheit der Längenmessung der Meter. Dieser ist definiert als die „Wegstrecke, die das Licht im Vakuum während einer Zeit von  $1/299\,792\,458$  Sekunde zurücklegt“ (Giancoli, 2010, S. 8). In der Psychologie gibt es keine vergleichbaren Festlegungen von Merkmalen und deren Maßeinheiten. Sie untersucht

Merkmale wie Persönlichkeitseigenschaften, deren Ausprägungen nicht direkt beobachtbar sind, sondern über beobachtbares Verhalten, beispielsweise in Form von Fragebogenantworten, erschlossen werden müssen. Auf Basis von Messmodellen können lediglich Schätzungen der entsprechenden Merkmalsausprägungen vorgenommen werden. Die Validität psychologischer Messungen kann nicht anhand eines Messinstrumentenvergleichs zuverlässig bestimmt werden, sondern muss mithilfe umfangreicher Studien untersucht werden (Eid & Schmidt, 2014).

*Psychometrie* ist die Teildisziplin der Psychologie, die sich mit den Problemen und Möglichkeiten psychologischer Messungen auseinandersetzt. Ihre Hauptaufgabe besteht darin, theoretische Ansätze für psychologische Messungen zu entwickeln und zu verbessern (Nunnally & Bernstein, 1994). Die Überprüfung angenommener Messmodelle ist allerdings schwierig aufgrund des bereits erwähnten Fehlens allgemein akzeptierter Merkmalsdefinitionen und Maßeinheiten (Eid & Schmidt, 2014).

Die eingangs gestellte Frage nach der Messbarkeit psychologischer Phänomene bedarf also einer Umformulierung: Misst die Psychologie tatsächlich das, was sie vorgibt zu messen? Um die Validität psychologischer Messungen und damit den praktischen Nutzen psychologischer Messmodelle zu demonstrieren, stellten Bortolotti, Tezza, Andrade, Bornia und Sousa Júnior (2013) der psychologischen Messung physikalische Messgrößen gegenüber. Sie konstruierten eine Skala mit 27 Items zur Erfassung der Körpergröße. Auf einer dichotomen Antwortskala sollten Personen Items wie „Do I think I would do well in a basketball team?“ (Bortolotti et al., 2013, S. 2 350) beantworten. Am Ende der Skala sollten sie ihre Körpergröße in Zentimetern angeben, die physikalische bzw. reale Körpergröße. Die aus der psychologischen Messung geschätzten Parameter für das Merkmal Körpergröße stellten sie dann der physikalischen bzw. realen Messgröße gegenüber. Dabei zeigte sich ein starker Zusammenhang zwischen den Parametern der geschätzten Körpergröße und den realen Werten (Bortolotti et al., 2013).

Van der Linden (2016a) zog ebenfalls eine Körpergrößenmessung mittels dichotomer Items (z. B. „I bump my head quite often.“; van der Linden, 2016a, S. 27) heran, um den Nutzen logistischer Item Response Modelle zu veranschaulichen (für einen Überblick zu Item Response Modellen vgl. Eid & Schmidt, 2014; Rost, 2004; van der Linden, 2016b). Laut van der Linden sei dieses Vorgehen ungewöhnlich, da Körpergröße eine physikalische Größe zu

sein scheint, die nur durch einen Meterstab gemessen werden könne. Jedoch wirkten sich Unterschiede in der Körpergröße auf das Verhalten aus, weshalb es durchaus möglich sei, diese Unterschiede als Indikatoren für Körpergröße zu betrachten und daraus eine psychologische Messung abzuleiten. Van der Linden bedauerte, dass er nicht die absolute Körpergröße mit erhob, resümierte aber "how much richer a plain variable as body height becomes when it is presented lined with behavioral consequences rather than a single number for each subject" (van der Linden, 2016a, S. 27f).

Hilbert und Kollegen (Hilbert, 2016; Hilbert et al., 2018) beschäftigten sich mit der Annahme, dass eine gelungene Messung die Relationen der empirischen Werte auf der Ratingskala abbilden müsste. Dabei untersuchten sie, bis zu welchem Grad Ratingskalen metrische Informationen wiedergeben können, d. h. wie gut Relationen wie Ordnung, Unterschiede und Verhältnisse der empirischen bzw. realen Werte bei der Abbildung erhalten bleiben. Ähnlich wie Bortolotti und Kollegen (2013) untersuchten sie empirisch den Zusammenhang zwischen realen Messgrößen (Körpergewicht, Körpergröße, Alter und Einkommen) und deren Abbildung auf Fragebogenitems. Nachdem sie die realen Werte der Merkmale Körpergröße und Körpergewicht sowie Einkommen und Alter abfragten, sollten sich die Probanden auf verschiedenen Skalentypen (fünf- und siebenstufige Skala, dichotome Skala und visuelle Analogskala) hinsichtlich dieser Messgrößen selbst einschätzen. Sie konnten zeigen, dass Fragebogenantworten reale Werte allenfalls in ordinaler Form widerspiegeln. Außerdem stellten sie fest, dass die Interpretierbarkeit von Fragebogenergebnissen durch Variablen beeinflusst wird, die den Referenzrahmen der Selbstbewertung bestimmen (z. B. Geschlecht).

Aufbauend auf diesen Ergebnissen und mit ähnlichem Vorgehen untersuchte Eichhorn (2016), ob die Perspektive, aus der sich eine Person selbst beurteilt (Selbstbewertungsperspektive), eine weitere den Referenzrahmen der Selbstbewertung bestimmende Variable ist und ob die Selbstbewertung durch die Darbietung merkmalsrelevanter statistischer Informationen beeinflusst wird. Im Fokus der Analysen stand erneut die Frage, unter welcher Bedingung die Relationen der realen Werte im Mittel am besten auf der Antwortskala erhalten bleiben. Dabei zeigte sich, dass die Selbstbewertungsperspektive beeinflusst, wie gut die Relationen der realen Werte auf der

Antwortskala abgebildet werden. Die Darbietung merkmalsrelevanter statistischer Informationen hatte nur bei Männern und nur bei bestimmten Merkmalen einen Einfluss.

Die vorliegende Arbeit setzt direkt die Arbeiten von Hilbert und Kollegen (Hilbert, 2016; Hilbert et al., 2018) sowie Eichhorn (2016) fort. Ihre übergeordnete Fragestellung ist, ob die Psychologie tatsächlich das misst, was sie vorgibt zu messen. In drei empirischen Studien werden verschiedene Aspekte der Gültigkeit psychologischer Messungen untersucht und welche Einflussfaktoren dabei eine Rolle spielen. Erneut werden der psychologischen Messung physikalische Messgrößen gegenübergestellt.

Die Studien I und II untersuchen zunächst die psychometrische Güte einzelner Items, die physikalisch messbare Personenmerkmale (Alter, Körpergröße, Körpergewicht und Einkommen) erfassen. Zentrale Fragestellungen sind, ob Bildungsniveau und relatives Körpergewicht (Studie I) sowie Itemformulierung und Skalenformat (Studie II) Einflussvariablen sind, die dazu führen, dass die Relationen der realen Werte im Mittel unterschiedlich gut auf der Antwortskala erhalten bleiben.

Studie III untersucht den Zusammenhang zwischen physikalisch messbaren Personenmerkmalen (Alter, Körpergröße und Körpergewicht) und deren Abbildung auf eigens entwickelten psychologischen Skalen. Die Konstruktion der Skalen folgt dabei den in der Fachliteratur gegebenen Empfehlungen zur Fragebogenkonstruktion. Zentrale Fragestellung ist, wie gut die Skalen die physikalischen Merkmale abbilden, genauer gesagt, wie gut der erstellte Fragebogen in der Lage ist, die Merkmale zu messen, die er messen soll.

## **2. Studie I: Einfluss von Bildungsniveau und relativem Körpergewicht auf das Skalenniveau**

### **2.1 Theorie Studie I**

#### **2.1.1 Einführung**

Studie I ist eine Reanalyse bestehender Daten, die im Rahmen der Masterarbeit der Autorin (Eichhorn, 2016) erhoben und analysiert wurden. Der Datensatz wurde inzwischen zur Veröffentlichung eingereicht (Hilbert et al., 2018). Im Rahmen der Masterarbeit wurde mithilfe eines zweifaktoriellen gemischten Designs untersucht, ob die Selbstbewertungsperspektive eine den Referenzrahmen der Selbstbewertung bestimmende Variable ist und ob die Selbstbewertung durch die Darbietung merkmalsrelevanter statistischer Informationen beeinflusst wird. Zunächst wurden die realen Werte körperbezogener Merkmale (Alter, Körpergröße, Schuhgröße und Körpergewicht) und das monatliche Nettoeinkommen erhoben. Anschließend bewerteten sich alle Teilnehmer ( $n = 2\,094$ ) in den zuvor erhobenen Merkmalen aus drei verschiedenen, durch unterschiedliche Itemformulierung induzierten Perspektiven auf einer fünfstufigen Likert-Skala (Likert, 1932). Dabei wurden Balkendiagramme mit Populations-Häufigkeitsverteilungen des jeweiligen Merkmals in vier Variationen eingeblendet. Untersucht wurde, unter welcher Bedingung die Relationen der realen Werte im Mittel am besten auf der Antwortskala erhalten bleiben. Es zeigte sich, dass die Selbstbewertung in Abhängigkeit vom Geschlecht bei allen Merkmalen unterschiedlich ausfiel. Die Selbstbewertungsperspektive zeigte sich bei den Variablen Alter und Körpergewicht (beide Geschlechter) sowie bei Schuhgröße (Frauen) und Einkommen (Männer) als eine weitere den Referenzrahmen der Selbstbewertung bestimmende Variable. Die Darbietung merkmalsrelevanter statistischer Informationen wirkte sich nur bei den Merkmalen Körpergewicht und Einkommen bei Männern aus.

Bei der Datenanalyse blieb das Bildungsniveau als möglicher Faktor, der die Selbstbewertung beeinflussen könnte, unbeachtet. Zudem warf die Diskussion der Ergebnisse die Frage auf, ob das um die Körpergröße korrigierte Körpergewicht mit einer höheren Prädiktionsgüte einhergeht als das absolute Körpergewicht (Eichhorn, 2016). Zur Klärung dieser Fragen wird im Rahmen der vorliegenden Arbeit eine Datenreanalyse durchgeführt. Im Mittelpunkt der Reanalyse steht erneut die Frage, unter welchen der Bedingungen sich im Mittel die größte Annäherung an Erhalt von Ordnung, Unterschieden

und Verhältnissen der realen Werte auf der Antwortskala ergibt. Der theoretische Hintergrund wird in verkürzter Form wiedergegeben (Abschnitte 2.1.2 und 2.1.3). Für eine detaillierte Übersicht wird auf Eichhorn (2016) verwiesen. Die Methodik der erneuten Analysen sowie deren Ergebnisse werden ausführlich berichtet. Die Diskussion bezieht sich auf die Ergebnisse der Reanalysen.

### **2.1.2 Der Referenzrahmen als Einflussgröße**

Die Wahrnehmungspsychologie unterscheidet bei der Raumwahrnehmung die allozentrische und die egozentrische Perspektive. Die allozentrische Perspektive ist die Wahrnehmung aus einer äußeren und unbeteiligten Perspektive. Die egozentrische Perspektive ist eine aktive Wahrnehmung des Raumes, bei der der Betrachter selbst eine zentrale Position einnimmt. Mit jedem Blickwechsel und jeder Bewegung verändert sich dabei der Raum (Brewer & Pears, 1999; Klatzky, 1998). Die kognitive Neuropsychologie vermutet im Gehirn ebenfalls allozentrische und egozentrische Referenzsysteme. Dabei werden neuronale Raumrepräsentationen beschrieben, die sich entweder auf einen äußeren, objektiven Raum oder auf das eigene Ich bzw. den eigenen Körperstandpunkt beziehen (Berthoz, 1991; Dieterich, 2012). Allgemein gefasst gilt jede Äußerung in einem objektiven Referenzrahmen als allozentrisch und jede Äußerung in einem subjektiven Referenzrahmen, dessen Mittelpunkt das sich äußernde Subjekt darstellt, als egozentrisch (Klatzky, 1998).

Die Sozialpsychologie überträgt die Konzepte des allozentrischen und egozentrischen Referenzrahmens auf das Selbst des Menschen (Triandis, Leung, Villareal & Clack, 1985). Allozentrisch orientierte Personen definieren sich in Bezug auf soziale Gruppen und erleben sich stärker kontextabhängig. Egozentrisch orientierte Menschen betonen mehr ihre Einzigartigkeit und Eigenständigkeit und sind weniger kontextbezogen (Triandis, 2001).

Die Diskrepanz zwischen verschiedenen Perspektiven findet sich auch im Fischteichereffekt wieder. Dabei wird das Phänomen beschrieben, dass das akademische Selbstkonzept einer Person vom Referenzrahmen abhängt und durch interindividuelle Standards bestimmt wird, gegen die Personen ihre schulischen Leistungen vergleichen. Bei ähnlich begabten Schülern haben diejenigen ein höheres akademisches Selbstkonzept, die eine Schule mit niedrigem Leistungsniveau besuchen, als diejenigen auf einer Schule mit hohem Leistungsniveau (Marsh, 1984; Marsh, 1987; Marsh & Parker, 1984).



In einer Studie konnte gezeigt werden, dass Amerikanerinnen asiatischer Abstammung signifikant besser bei Mathematikaufgaben abschneiden, nachdem ihr asiatisches Identitätsgefühl aktiviert wurde, als nachdem ihr weibliches Identitätsgefühl aktiviert wurde (Shih, Pittinsky & Ambady, 1999). Verschiedene Aspekte des Selbst lassen sich in Abhängigkeit von Stimmung, Zeitpunkt und sozialem Kontext aktivieren (Morf & Koole, 2014). Dabei treten physische Merkmale, Fähigkeiten, Rollen, Verhalten, Persönlichkeitseigenschaften oder Einstellungen abhängig vom jeweiligen Referenzrahmen unterschiedlich stark hervor (Simon, 2012). Dies führt dazu, dass menschliches Erleben und Verhalten abhängig von der zugrundeliegenden Selbstkonzeption und dem gewählten Referenzrahmen unterschiedlich ausfällt, obwohl Menschen dazu neigen, ihr Selbst als stabile Einheit und als etwas zutiefst Persönliches und weitestgehend Kontextunabhängiges wahrzunehmen (Morf & Koole, 2014).

Für die Fragebogenkonstruktion hat dies zur Konsequenz, dass je nach Referenzrahmen, innerhalb dessen Informationen erhoben werden, die Antwortkategorien einer Ratingskala andere reale Werte abbilden können. Bleibt dies unberücksichtigt, könnte es sein, dass die Antwortskala nicht einmal die Rangordnung der realen Werte abbildet und damit selbst den Kriterien einer Ordinalskala nicht mehr genügt (Hilbert, 2016; Hilbert et al., 2018).

### **2.1.3 Die graphische Darstellung statistischer Informationen**

Die Aufbereitung und Darstellung von Daten weist eine über tausend Jahre alte Geschichte auf. Dokumentiert und weitergeführt wird diese von Friendly und Denis (2001) in ihrem Projekt *Milestones in the history of thematic cartography, statistical graphics, and data visualization*. Playfair (1786) revolutionierte die Visualisierung statistischer Informationen. Vor mehr als 200 Jahren erfand er Torten-, Balken- und Zeitreihendiagramme, die bis heute nahezu unverändert eingesetzt werden (Bissantz & Butterwegge, 2013).

Graphiken haben keine ausschließlich illustrative Funktion, sondern sollen die Aufnahme und Kommunikation statistischer Informationen vereinfachen (Weber & Wenzel, 2013). Dabei beschäftigt sich umfangreiche Forschung mit der Frage, was das Verständnis graphischer Darstellungen beeinflusst. Auf der Seite des Graphen werden Inhalt, Komplexität, Typ und optische Charakteristik als Einflussfaktoren betrachtet. In Bezug auf den Rezipienten werden mathematisches Wissen, domänenspezifisches Vorwissen,

Intelligenz und Vertrautheit im Umgang mit Graphen als Einflussgrößen diskutiert (vgl. Curcio, 1987; Friel, Curcio & Bright, 2001; Harvey & Bolger, 1996; Shah, 2002; für Reviews vgl. Glazer, 2011; Shah & Hoeffner, 2002).

Die am häufigsten verwendeten graphischen Darstellungen sind Balkendiagramme, weil sie am leichtesten zu verstehen und zu lesen sind. Bereits fünfjährige Kinder sind in der Lage, sich Informationen aus Balkendiagrammen zu erschließen (Russell, Schifter & Bastable, 2002). Im Vergleich zu anderen Diagrammtypen weisen sie die geringste Komplexität auf und benötigen zu ihrer Interpretation und ihrem Verständnis kaum Übung und Domänenwissen (Culbertson & Powers, 1959; Glazer, 2011; Shah, 2002). Statistische Informationen transportieren sie unter allen Graphen mit den geringsten Verzerrungseffekten (Shah, 2002).

Graphische Darstellungen weisen zahlreiche und gut etablierte Kommunikationsvorteile auf. Sie nutzen die Vorzüge der visuellen Wahrnehmung, bei der viele Informationen unmittelbar und gleichzeitig verarbeitet werden können (Shah, 2002). Durch die bildliche Darstellung werden der Vergleich und das Erkennen von Trends, Mustern und Abweichungen vereinfacht (Harris, 1999). Die graphische Darstellung statistischer Informationen hat einen maßgeblichen Vorteil: Sie kann Erkenntnis unmittelbar veranschaulichen und sichtbar machen, wohingegen der sprachlichen Darstellung erst eine Verkettung logischer Verknüpfungen folgen muss (Weber & Wenzel, 2013).

#### **2.1.4 Rationale**

Wie Hilbert und Kollegen (Hilbert, 2016; Hilbert et al., 2018) beschäftigte sich Eichhorn (2016) mit der Annahme, dass eine gelungene Messung die Relationen der empirischen Werte auf der Ratingskala abbilden müsste (vgl. Abschnitt 1.). Untersucht wurde, bis zu welchem Grad Ratingskalen metrische Informationen wiedergeben können, d. h. unter welcher Bedingung sich im Mittel die größte Annäherung an Erhalt von Ordnung, Unterschieden und Verhältnissen der realen Werte auf der Antwortskala ergibt. Für die Realisierung wurde ein zweifaktorielles gemischtes Design verwendet. Zunächst wurden von allen Teilnehmern Alter, Körpergröße, Körpergewicht, Schuhgröße und monatliches Nettoeinkommen erhoben. Die ermittelten Werte gingen als die realen Werte der untersuchten Merkmale in die Datenanalyse ein.

*Faktor mit Messwiederholung.* Es wurde angenommen, dass die Selbstbewertungsperspektive eines Befragten eine weitere den Referenzrahmen der Selbstbewertung bestimmende Variable ist. Deswegen sollten sich alle Teilnehmer aus drei verschiedenen Perspektiven in den zuvor erhobenen Merkmalen (Alter, Körpergröße, Schuhgröße, Körpergewicht und monatliches Nettoeinkommen) auf einer fünfstufigen Likert-Skala (Likert, 1932) selbst bewerten. Dabei wurden in Form eines explorativen Vorgehens die Konzepte des allozentrischen und egozentrischen Referenzrahmens auf die Itemformulierung angewendet. Bei der Testperson sollte ein innerer Perspektivenwechsel durch drei unterschiedliche Formulierungen (z. B. „Ich bin alt.“; „Ich halte mich für alt.“; „Andere halten mich für alt.“) erreicht werden (vgl. Klatzky, 1998; Mandel, 2003).

*Faktor ohne Messwiederholung.* Es wurde postuliert, dass die Darbietung merkmalsrelevanter statistischer Informationen eine Hilfestellung bei der Selbstbewertung ist, was in einer genaueren Selbsteinschätzung resultieren sollte. Deshalb wurden Balkendiagramme, die die Häufigkeitsverteilungen der zu bewertenden Merkmale in der Population abbilden, während der Selbstbewertung in den folgenden vier Varianten angezeigt: Gesamtmerkmalsverteilungen in der Population (GV), geschlechtsspezifische Merkmalsverteilungen in der Population (GSV), Gesamtmerkmalsverteilungen und geschlechtsspezifische Merkmalsverteilungen in der Population (GV/GSV) sowie Anzeige keiner Verteilung (KV) als Kontrollbedingung.

Für die Reanalyse der Daten wird das Merkmal Schuhgröße aufgrund eingeschränkter Varianz und zu grobem Maßstab ausgeschlossen. Die Analysen des Merkmals Einkommen werden mit skalierten Werten durchgeführt (Tausend-Euro-Werte). Es werden dadurch besser interpretierbare Regressionskoeffizienten erwartet.

Bei der ersten Datenanalyse blieb das Bildungsniveau als möglicher Faktor, der die Selbstbewertung beeinflussen könnte, unbeachtet. Wie unter Abschnitt 2.1.3 dargestellt, gibt es zahlreiche Faktoren, wie beispielsweise Intelligenz und mathematisches Wissen, die das Verständnis graphischer Darstellungen beeinflussen können (z. B. Glazer, 2011; Shah & Hoeffner, 2002). Möglicherweise können Menschen mit höherem Bildungsniveau Graphiken besser lesen und verstehen, weshalb sie die von den Graphiken transportierten Informationen eher in ihre Entscheidungen integrieren können als Personen mit niedrigerem Bildungsniveau. Im Zuge der Datenreanalyse wird diese Frage untersucht.

Die erste Datenanalyse zeigte, dass das absolute Körpergewicht ein Prädiktor für das Antwortverhalten bei Selbsteinschätzungen zum eigenen Gewicht ist (Eichhorn, 2016). Dennoch schließt dies nicht aus, dass sich Personen mit demselben Gewicht und stark divergierenden Körpergrößen unterschiedlich bewerten. Aus diesem Grund wird bei der Reanalyse zur Untersuchung der Beziehung zwischen den realen Werten und der Antwortskala der Body Mass Index (BMI) herangezogen. Der BMI ist ein um die Körpergröße korrigiertes Maß des Körperwichts, definiert als Körpergewicht in Kilogramm dividiert durch die quadrierte Körpergröße in Metern (Birbaumer & Schmidt, 2010). Durch Regressionsanalysen auf Basis des BMI soll überprüft werden, ob der BMI einen inkrementellen Beitrag für die Vorhersage des Antwortverhaltens bei Selbsteinschätzungen des Gewichts über das Körpergewicht hinaus liefert.

## **2.2 Methoden Studie I**

### **2.2.1 Stichprobe**

Die Stichprobe umfasste 2 094 Teilnehmer im Alter von 18 bis 95 Jahren ( $M = 28$ ,  $SD = 11.2$ ), darunter 1 636 Frauen (78.1 %) und 458 Männer (21.9 %). Das Bildungsniveau verteilte sich wie folgt: 0.9 % Kein Schulabschluss, 11.1 % Hauptschulabschluss, 29.6 % Mittlere Reife, 29.4 % Fachabitur/Abitur, 29.0 % Hochschulabschluss/Universitätsabschluss. Die Gesamtstichprobe setzte sich aus vier unabhängigen Substichproben zusammen, die sich auf die Bedingungen GV, GSV und GV/GSV sowie die Kontrollbedingung KV verteilten. Tabelle 2.2.1. zeigt eine Übersicht der Stichprobenzusammensetzung. Dabei wurde das in fünf Kategorien erhobene Bildungsniveau in drei Kategorien zusammengefasst (siehe Methodenteil, Abschnitt 2.2.4).

Die Anwerbung der Teilnehmer erfolgte über private E-Mail-Verteiler der Autorin und über das soziale Netzwerk Facebook. Die Teilnahme war freiwillig und ohne Vergütung.

Tabelle 2.2.1.

Verteilung von Geschlecht, Bildungsniveau und Alter innerhalb der Bedingungen GV<sup>1</sup>, GSV<sup>2</sup>, GV/GSV<sup>3</sup> und KV<sup>4</sup>

Merkmal	GV <sup>1</sup>		GSV <sup>2</sup>		GV/GSV <sup>3</sup>		KV <sup>4</sup>	
	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%
Geschlecht								
Frauen	411	77.5	412	77.9	411	78.7	402	78.4
Männer	119	22.5	117	22.1	111	21.3	111	21.6
Bildungsniveau								
Kein Abitur	223	42.1	223	42.2	225	43.1	200	39.0
Abitur	166	31.3	143	27.0	140	26.8	167	32.6
Hochschulabschluss	141	26.6	163	30.8	157	30.1	146	28.5
	<i>Md</i>	<i>SD</i>	<i>Md</i>	<i>SD</i>	<i>Md</i>	<i>SD</i>	<i>Md</i>	<i>SD</i>
Alter	28	11.7	28	10.8	29	11.0	27	11.3
Gesamtteilnehmerzahl	530		529		522		513	

**Anmerkungen.**<sup>1</sup> Anzeige Balkendiagramm mit der Gesamtmerkmalsverteilung in der Population.<sup>2</sup> Anzeige Balkendiagramm mit der geschlechtsspezifischen Merkmalsverteilung in der Population.<sup>3</sup> Gleichzeitige Anzeige Balkendiagramme mit der Gesamtmerkmalsverteilung und der geschlechtsspezifischen Merkmalsverteilung in der Population.<sup>4</sup> Anzeige keiner Merkmalsverteilung. Bei der Bedingung KV handelt es sich um die Kontrollbedingung.*n* = Anzahl; % = prozentualer Anteil innerhalb der Teilstichprobe; *Md* = Median; *SD* = Standardabweichung.**2.2.2 Material**

Für die Studie wurde ein Fragebogen erstellt, in dem zuerst Alter, Körpergröße, Körpergewicht, Schuhgröße und monatliches Nettoeinkommen der Teilnehmer (numerisch und ganzzahlig) erfragt wurden. Diese Daten gingen als die realen Werte der untersuchten Merkmale in die Analyse ein. Davon sind die wahren Werte eines Merkmals im Sinne der Klassischen Testtheorie (KTT) zu unterscheiden. Diese werden darin definiert als Erwartungswerte unendlich oft wiederholter und unabhängiger Messungen an derselben Person (Bühner, 2011). In dieser Arbeit wird nicht Bezug genommen auf die Definition der KTT, weshalb eine verbale Abgrenzung vorgenommen wurde. Die untersuchten Merkmale sind mindestens intervallskaliert.

Überdies wurden der höchste Bildungsabschluss (*Kein Schulabschluss, Hauptschulabschluss, Mittlere Reife, Fachabitur/Abitur oder Hochschulabschluss/Universitätsabschluss*) sowie das Geschlecht (*weiblich oder männlich*) erfragt.

Im Folgenden wurden die Teilnehmer aufgefordert, sich bezüglich der abgefragten fünf Merkmale ohne implizite Perspektive (z. B. „Ich bin alt.“), aus egozentrischer (z. B. „Ich halte mich für alt.“) und aus allozentrischer Perspektive (z. B. „Andere halten mich für alt.“) auf einer fünfstufigen Likert-Skala (Likert, 1932) selbst zu bewerten. Die verbale Verankerung der Antwortkategorien war *völlig unzutreffend*, *unzutreffend*, *weder noch*, *zutreffend* und *völlig zutreffend*. Die numerische Verankerung erfolgte von 1 (*völlig unzutreffend*) bis 5 (*völlig zutreffend*). Die gewählte Antwortkategorie stellte den beobachteten Wert der Person des jeweiligen Merkmals dar. Eine Übersicht aller Items findet sich in Anhang A (Tabelle A.1.; entnommen aus Eichhorn, 2016).

Zur graphischen Darstellung der Häufigkeitsverteilungen der abgefragten fünf Merkmale wurden Balkendiagramme erstellt: Geschlechtsspezifische Verteilungen für Frauen (rote Balken) und Männer (blaue Balken) sowie Gesamtverteilungen für beide Geschlechter (grüne Balken). Auch wenn bei der Darstellung metrisch-stetiger Daten Histogramme vorzuziehen sind (Büchter & Henn, 2007), wurden Balkendiagramme aufgrund besserer Verständlichkeit und Lesbarkeit verwendet. Dafür mussten die Daten in Klassen eingeteilt werden (z. B. bei Gewicht 50 – 59 kg), wobei Klassenbreite und -anzahl mit dem Ziel einer möglichst leichten Verständlichkeit gewählt wurden. Als Datenbasis dienten Zahlen aus dem deutschen Fußreport 2009 (Richter & Schaefer, 2009) und Daten des GESIS - Leibniz-Institut für Sozialwissenschaften (2015). Beide Datensätze bezogen sich auf in Deutschland lebende Erwachsene ab 18 Jahren. Eine vollständige Übersicht aller verwendeten Balkendiagramme findet sich in Anhang A (Abbildungen A.1. bis A.5.)

Der Einfluss von visuell dargebotenen Merkmalsverteilungen auf die Selbstbewertung wurde mithilfe der folgenden vier Bedingungen ohne Messwiederholung untersucht: Unter Bedingung GV wurde oberhalb der jeweiligen Fragebogenitems ein Balkendiagramm mit der Populationsgesamtverteilung des abgefragten Merkmals angezeigt, unter Bedingung GSV ein Balkendiagramm mit der geschlechtsspezifischen Verteilung des abgefragten Merkmals und unter Bedingung GV/GSV wurden zwei Balkendiagramme (Populationsgesamtverteilung und geschlechtsspezifische Verteilung) angezeigt. Unter Bedingung KV wurde den Teilnehmern der Fragebogen ohne Balkendiagramme dargeboten.

### 2.2.3 Erhebung

Die Datenerhebung erfolgte vom 16. Mai bis 19. Juni 2016 mittels eines Onlinefragebogens. Dieser wurde mit der Befragungssoftware SoSci Survey erstellt (Leiner, 2014) und den Teilnehmern auf [www.soscisurvey.de](http://www.soscisurvey.de) zur Verfügung gestellt. Eine Kontrolle äußerer Bedingungen war nicht möglich. Für die Teilnahme wurde ein Mindestalter von 18 Jahren festgelegt, da die eingeblendeten Merkmalsverteilungen auf Daten volljähriger Erwachsener basierten und sich jede Person darin wiederfinden können sollte. Darüber hinaus gab es keine weiteren Ein- oder Ausschlusskriterien.

Vor der Befragung klärte eine Textseite die Teilnehmer über den Zweck der Untersuchung, die Art der erhobenen Daten sowie über Freiwilligkeit und Anonymität der Teilnahme auf. Mittels Klicken des „Weiter“-Buttons gaben die Befragten ihr Einverständnis mit den Durchführungsbedingungen.

Auf der folgenden Seite wurden zunächst die Merkmale Alter, Bildungsniveau, Körpergröße, Schuhgröße, Körpergewicht, monatliches Nettoeinkommen und Geschlecht aller Teilnehmer erhoben. Nach Abschluss der ersten Fragebogenseite erfolgte mittels einer Urnenfunktion ohne Zurücklegen die zufallsgesteuerte Zuweisung der Probanden zu einer der vier Bedingungen ohne Messwiederholung. Die Urnenfunktion aktivierte sich automatisch, sobald der Teilnehmer die erste Seite vollständig bearbeitet und den „Weiter“-Button geklickt hatte. Sie stellte sicher, dass Experimental- und Kontrollgruppen annähernd gleich groß ausfielen (Bosch, 2011). In der Urne befanden sich die Fragebogenversionen der vier Bedingungen ohne Messwiederholung. Die Balkendiagramme wurden entsprechend dem ausgewählten Geschlecht angezeigt.

Auf den weiteren fünf Fragebogenseiten erfolgte jeweils die Selbstbewertung aus drei verschiedenen Perspektiven (je nach Bedingung mit oder ohne Anzeige des Balkendiagramms bzw. der Balkendiagramme), wobei auf jeder Seite genau ein Merkmal zu bewerten war. Das Fragebogenlayout wurde so optimiert, dass bei üblichen Bildschirmauflösungen kein Scrollen notwendig war. Dies stellte sicher, dass die entsprechenden Balkendiagramme während der Selbstbewertung für die Probanden vollständig sichtbar waren.

Abschließend wurde die erfolgreiche Speicherung der Daten bestätigt und den Probanden für ihre Teilnahme gedankt. Ein vollständiger Onlinefragebogen (Version

GV/GSV-Frauen) findet sich in Form von Screenshots in Anhang A (Abbildungen A.6. bis A.13.)

#### 2.2.4 Analysen

Die Datenreanalyse erfolgte mit der Open Source Statistiksoftware R (R Core Team, 2017) unter Verwendung der R-Pakete „foreign“ (R Core Team, 2016), „dplyr“ (Wickham & Francois, 2016) und „ordinal“ (Christensen, 2015). Die Diagramme wurden mit dem R-Paket „ggplot2“ (Wickham, 2009) erstellt.

Bei der Datenreanalyse wurden, wie bereits in Eichhorn (2016), beide Geschlechter getrennt voneinander analysiert, da die Ergebnisse in Hilbert (2016; Hilbert et al., 2018) geschlechtsspezifische Unterschiede erwarten ließen. Die Reanalyse wurde nur für die Merkmale Alter, Körpergröße, Körpergewicht und Einkommen durchgeführt. Das Merkmal Schuhgröße wurde aufgrund eingeschränkter Varianz und zu grobem Maßstab ausgeschlossen. Beim Merkmal Einkommen wurden die Regressionsanalysen mit skalierten Werten, nämlich Tausend-Euro-Werten (kEUR), durchgeführt. Dabei wurde das angegebene monatliche Nettoeinkommen jedes Teilnehmers durch tausend dividiert. Alle Datenreanalysen wurden mit der Funktion *c/m* durchgeführt, da sie bei mehr als einem Prädiktor stabilere Schätzungen liefert als die Funktion *c/m2* (Christensen, 2015). In Eichhorn (2016) wurde die Funktion *c/m2* verwendet. Bei nur einem Prädiktor (z. B. Körpergewicht) fallen die Schätzungen für beide Funktionen identisch aus (Christensen, 2015), weswegen ein Vergleich mit den Ergebnissen aus Eichhorn (2016) unproblematisch ist.

Die Teilnehmer gaben auf einer fünfstufigen Ratingskala den Grad ihrer Zustimmung zu den jeweiligen Itemaussagen an. Die fünf Antwortalternativen wurden wie folgt kodiert: *völlig unzutreffend* mit der Ziffer 1, *unzutreffend* mit der Ziffer 2, *weder noch* mit der Ziffer 3, *zutreffend* mit der Ziffer 4 und *völlig zutreffend* mit der Ziffer 5.

Die Daten wurden unter Berücksichtigung des Bildungsniveaus deskriptiv-statistisch ausgewertet. Dabei wurden Mittelwert, Median und Standardabweichung für die Merkmale Alter, Körpergröße, Körpergewicht und Einkommen innerhalb jeder Bedingung (GV, GSV, GV/GSV und KV) getrennt nach Bildungsniveau separat für Frauen und Männer ermittelt und tabellarisch gegenübergestellt.

*Bildungsniveau.* Für die Merkmale Alter, Körpergröße, Körpergewicht und Einkommen wurde überprüft, ob das Bildungsniveau ein weiterer die Selbstbewertung bestimmender



Faktor ist. Da einige Bildungsabschlüsse nur selten beobachtet wurden (z. B. *Kein Schulabschluss*), wurden zur einfacheren Interpretation die Kategorien *Kein Schulabschluss*, *Hauptschulabschluss* und *Mittlere Reife* zur Kategorie *Kein Abitur* zusammengefasst. Die Kategorien *Fachabitur/Abitur* und *Hochschulabschluss/Universitätsabschluss* blieben inhaltlich unverändert, wurden aber in *Abitur* bzw. *Hochschulabschluss* umbenannt. Die Kategorie *Abitur* wurde als Referenzgruppe kodiert. Mithilfe von Proportional Odds Regressionen wurde für alle Merkmale (Alter, Körpergewicht (BMI), Körpergröße und Einkommen) untersucht, ob sich die Stärke des Zusammenhangs zwischen den realen Werten und der Antwortskala in Abhängigkeit vom Bildungsniveau signifikant verändert. Dabei wurde ein flexibles Modell ohne äquidistante Schwellen gewählt, da es sich als das generell angemessenere Modell zeigte (Eichhorn, 2016). Bei der Schätzung komplexerer Modelle wird eine Zentrierung der Variablen (Alter, Körpergröße, Körpergewicht und Einkommen) auf den Stichprobenmittelwert empfohlen. Eine Schätzung mit unzentrierten Variablen führt zu unzuverlässigen Schätzungen der Standardfehler und deshalb zu weniger stabilen Ergebnissen (Christensen, 2015). Da die Hinzunahme des Bildungsniveaus als möglicher Faktor zu einem komplexeren Modell führt, wurden die Analysen mit auf den Stichprobenmittelwert zentrierten Werten durchgeführt.

*Körpergewicht.* Beim Merkmal Körpergewicht wurde die Beziehung zwischen den realen Werten und der Antwortskala mithilfe von Proportional Odds Regressionen untersucht. Die Regressionsanalysen wurden auf Basis des BMI durchgeführt. Dabei wurde das Körpergewicht jedes Teilnehmers in Kilogramm durch seine quadrierte Körpergröße in Metern dividiert (Birbaumer & Schmidt, 2010). Für jede Bedingung wurde ein Diagramm mit der Lage der Schwellenparameter (rot) sowie des Mittelwerts (blau) und des Medians (grün) der jeweiligen Teilstichprobe des nicht-messwiederholten Faktors erstellt. Mithilfe der Diagramme wurde überprüft, inwieweit die Abstände der Schwellenparameter zwischen den Bedingungen variieren und ob der Mittelwert der realen Werte in der Mittelkategorie der Skala liegt. Je besser die Informationen der realen Werte auf der Antwortskala erhalten bleiben, desto eher sollte sich eine Annäherung an äquidistante Schwellen ergeben und der Mittelwert der realen Werte in der Mittelkategorie liegen. Die Schwellenparameter mit ihren dazugehörigen Steigungskoeffizienten wurden tabellarisch berichtet.

Ergänzend wurde bei jeder Proportional Odds Regression ein Likelihood Ratio Test durchgeführt, der die Annahme äquidistanter Schwellen überprüft. Dabei wird ein weniger restriktives Modell mit flexiblen Schwellen gegen ein restriktiveres Modell mit äquidistanten Schwellen getestet. Ein signifikanter Likelihood Ratio Test zeigt an, dass das weniger restriktive Modell die Daten signifikant besser abbildet und die Annahme äquidistanter Schwellen verworfen werden muss (Norušis, 2012).

Die Analysen auf Basis des BMI wurden sodann verglichen mit den Analysen auf Basis des absoluten Körpergewichts in Kilogramm aus Eichhorn (2016). Das Signifikanzniveau wurde bei allen Tests auf  $\alpha = 5\%$  festgelegt. Eine klassische Hypothesenprüfung stand nicht im Fokus der Analysen, sondern eine Interpretation des Gesamtbildes der Daten mithilfe eines explorativen Vorgehens. Auf eine Adjustierung des  $\alpha$ -Fehlers wurde deshalb verzichtet. Alle geschätzten Werte bzw. Parameter waren signifikant von Null verschieden, falls nicht anders angegeben. Alle Signifikanztests wurden zweiseitig durchgeführt.

## **2.3 Ergebnisse Studie I**

### **2.3.1 Deskriptive Statistik**

Die Tabellen 2.3.1., 2.3.2., 2.3.3. und 2.3.4. zeigen Mittelwert, Standardabweichung und Median für alle Merkmale und Bedingungen getrennt nach Bildungsniveau und Geschlecht. Dabei ist ersichtlich, dass die Lagemaße der zentralen Tendenz beim Merkmal Alter über alle Bedingungen beim Bildungsniveau *Abitur* insgesamt am geringsten ausfallen (beide Geschlechter) und beim Merkmal Körpergewicht über alle Bedingungen beim Bildungsniveau *Kein Abitur* insgesamt am höchsten ausfallen (Frauen). Beim Merkmal Einkommen ergeben sich für beide Geschlechter über alle Bedingungen hinweg beim Bildungsniveau *Abitur* die niedrigsten Lagemaße der zentralen Tendenz und beim Bildungsniveau *Hochschulabschluss* tendenziell die höchsten. Dabei fallen Mittelwert und Median für Frauen immer geringer aus als für Männer. Beim Merkmal Körpergröße zeigen sich keine auffälligen Unterschiede in den Lagemaßen der zentralen Tendenz zwischen den Bedingungen.

Tabelle 2.3.1.

Mittelwert (*M*), Standardabweichung (*SD*) und Median (*Md*) der realen Werte getrennt nach Bildungsniveau und Geschlecht – Bedingung GV<sup>1</sup>

Merkmal	Kein Abitur				Abitur				Hochschulabschluss			
	<i>n</i>	<i>M</i>	<i>SD</i>	<i>Md</i>	<i>n</i>	<i>M</i>	<i>SD</i>	<i>Md</i>	<i>n</i>	<i>M</i>	<i>SD</i>	<i>Md</i>
Alter												
Frauen	173	35	12.4	33	135	29	9.2	25	103	32	9.8	28
Männer	50	33	15.2	29	31	29	13.1	25	38	36	11.8	32
Körpergröße												
Frauen	173	167	6.2	167	135	167	6.8	167	103	167	6.7	168
Männer	50	180	6.6	178.5	31	181	8.3	181	38	180	6.1	180
Körpergewicht												
Frauen	173	76	20.5	72	135	67	14.9	63	103	66	13.0	63
Männer	50	81	12.6	80	31	83	15.8	82	38	87	17.8	81
Einkommen												
Frauen	173	1 325	731	1 250	135	1 102	876	1 000	103	1 631	1 352	1 400
Männer	50	1 727	1 205	1 800	31	1 479	1 091	1 100	38	2 486	1 966	2 075

Anmerkungen.

*n* = Anzahl; *M* = Mittelwert; *SD* = Standardabweichung; *Md* = Median.

<sup>1</sup>Anzeige Balkendiagramm mit der Gesamtmerkmalsverteilung in der Population.

Tabelle 2.3.2.

Mittelwert (*M*), Standardabweichung (*SD*) und Median (*Md*) der realen Werte getrennt nach Bildungsniveau und Geschlecht – Bedingung GSV<sup>1</sup>

Merkmal	Kein Abitur				Abitur				Hochschulabschluss			
	<i>n</i>	<i>M</i>	<i>SD</i>	<i>Md</i>	<i>n</i>	<i>M</i>	<i>SD</i>	<i>Md</i>	<i>n</i>	<i>M</i>	<i>SD</i>	<i>Md</i>
Alter												
Frauen	171	34	12.1	31	116	27	8.3	25	125	32	8.5	29
Männer	52	35	12.9	32	27	27	10.0	23	38	34	10.7	30
Körpergröße												
Frauen	171	167	6.6	167	116	168	6.6	168.5	125	168	5.7	169
Männer	52	182	6.9	180	27	180	7.6	180	38	181	7.1	181.5
Körpergewicht												
Frauen	171	74	20.2	70	116	71	15.7	68	125	66	12.3	65
Männer	52	88	15.8	85.5	27	81	18.7	75	38	82	15.9	80
Einkommen												
Frauen	171	1 242	957	1 200	116	962	794	784	125	1 613	1 630	1 300
Männer	52	1 798	982	1 800	27	1 247	943	1 000	38	2 292	1 452	2 000

Anmerkungen.

*n* = Anzahl; *M* = Mittelwert; *SD* = Standardabweichung; *Md* = Median.

<sup>1</sup>Anzeige Balkendiagramm mit der geschlechtsspezifischen Merkmalsverteilung in der Population.

Tabelle 2.3.3.

Mittelwert (*M*), Standardabweichung (*SD*) und Median (*Md*) der realen Werte getrennt nach Bildungsniveau und Geschlecht – Bedingung GV/GSV<sup>1</sup>

Merkmal	Kein Abitur				Abitur				Hochschulabschluss			
	<i>n</i>	<i>M</i>	<i>SD</i>	<i>Md</i>	<i>n</i>	<i>M</i>	<i>SD</i>	<i>Md</i>	<i>n</i>	<i>M</i>	<i>SD</i>	<i>Md</i>
Alter												
Frauen	180	34	11.6	30.5	106	28	9.1	24	125	31	8.0	29
Männer	45	33	13.9	30	34	34	10.7	31.5	32	37	15.7	29
Körpergröße												
Frauen	180	167	6.2	167.5	106	168	6.9	168	125	168	6.3	169
Männer	45	179	6.0	180	34	183	8.0	182.5	32	179	7.3	180
Körpergewicht												
Frauen	180	75	18.5	72.5	106	71	19.3	65	125	66	14.9	62
Männer	45	85	20.0	84	34	89	15.5	87.5	32	82	21.9	74.5
Einkommen												
Frauen	180	1 278	753	1 200	106	1 155	1 263	832	125	1 797	2 604	1 385
Männer	45	1 849	1 292	1 800	34	1 959	1 313	1 950	32	2 241	1 673	1 900

Anmerkungen.

*n* = Anzahl; *M* = Mittelwert; *SD* = Standardabweichung; *Md* = Median.

<sup>1</sup> Gleichzeitige Anzeige Balkendiagramme mit der Gesamtmerkmalsverteilung und der geschlechtsspezifischen Merkmalsverteilung in der Population.

Tabelle 2.3.4.

Mittelwert (*M*), Standardabweichung (*SD*) und Median (*Md*) der realen Werte getrennt nach Bildungsniveau und Geschlecht – Bedingung KV<sup>1</sup>

Merkmal	Kein Abitur				Abitur				Hochschulabschluss			
	<i>n</i>	<i>M</i>	<i>SD</i>	<i>Md</i>	<i>n</i>	<i>M</i>	<i>SD</i>	<i>Md</i>	<i>n</i>	<i>M</i>	<i>SD</i>	<i>Md</i>
Alter												
Frauen	162	32	10.8	31	131	27	8.9	24	109	31	9.6	27
Männer	38	34	14.7	29	36	31	10.9	25	37	39	15.8	35
Körpergröße												
Frauen	162	168	6.1	168	131	168	6.6	168	109	169	6.4	169
Männer	38	180	6.1	179.5	36	182	6.7	182	37	180	5.8	180
Körpergewicht												
Frauen	162	76	18.8	70	131	68	17.3	63	109	67	14.6	64
Männer	38	87	18.1	84	36	86	18.5	81.5	37	80	13.5	78
Einkommen												
Frauen	162	1 270	768	1 200	131	1 069	993	750	109	1 433	1 215	1 200
Männer	38	1 890	956	1 974	36	1 774	1 548	1 400	37	3 065	4 119	2 100

Anmerkungen.

*n* = Anzahl; *M* = Mittelwert; *SD* = Standardabweichung; *Md* = Median.

<sup>1</sup> Anzeige keiner Merkmalsverteilung. Bei der Bedingung KV handelt es sich um die Kontrollbedingung.

### 2.3.2 Regressionsanalyse

*Bildungsniveau.* Die Tabellen 2.3.5., 2.3.6., 2.3.7. und 2.3.8. zeigen für die Bedingungen GV, GSV, GV/GSV und KV getrennt nach Geschlecht die Ergebnisse der Proportional Odds Regressionen (mit Bildungsniveau als zusätzlichen Prädiktor für die Stärke des Zusammenhangs zwischen den realen Werten und der Antwortskala) für die Merkmale Alter, Körpergewicht (BMI), Körpergröße und Einkommen.

Beim Merkmal Alter ist bei Frauen ein genereller, wenn auch uneindeutiger Zusammenhang zwischen den realen Werten und der Antwortskala ersichtlich (unter neun von zwölf Bedingungen signifikante Effekte für den Prädiktor *reale Werte* in der Referenzgruppe *Abitur*). Unter keiner Bedingung ist ein klares Muster erkennbar, dass sich die Stärke des Zusammenhangs in Abhängigkeit des Bildungsniveaus verändert. Bei Männern zeigt sich nur in der Fragenformulierung ohne implizite Perspektive ein Zusammenhang zwischen den realen Werten und der Antwortskala (unter drei von vier Bedingungen signifikante Effekte für den Prädiktor *reale Werte* in der Referenzgruppe *Abitur*). Überdies gibt es kein eindeutiges Anzeichen dafür, dass sich die Stärke des Zusammenhangs in Abhängigkeit des Bildungsniveaus verändert.

Bei den Merkmalen Körpergröße und Körpergewicht zeigt sich bei beiden Geschlechtern ein Zusammenhang zwischen den realen Werten und der Antwortskala (signifikante Effekte für den Prädiktor *reale Werte* in der Referenzgruppe *Abitur* über alle Bedingungen). Bei beiden Merkmalen gibt es weder bei Frauen noch bei Männern ein eindeutiges Anzeichen dafür, dass sich die Stärke des Zusammenhangs in Abhängigkeit des Bildungsniveaus verändert.

Beim Merkmal Einkommen zeigt sich bei Frauen ein Zusammenhang zwischen den realen Werten und der Antwortskala (signifikante Effekte für den Prädiktor *reale Werte* in der Referenzgruppe *Abitur* über alle Bedingungen). Unter keiner Bedingung ist ein deutliches Muster erkennbar, dass sich die Stärke des Zusammenhangs in Abhängigkeit des Bildungsniveaus verändert. Bei Männern ist ebenfalls ein Zusammenhang zwischen den realen Werten und der Antwortskala ersichtlich (unter elf von zwölf Bedingungen signifikante Effekte für den Prädiktor *reale Werte* in der Referenzgruppe *Abitur*). Es gibt ebenfalls keine eindeutigen Hinweise darauf, dass sich die Stärke des Zusammenhangs in Abhängigkeit des Bildungsniveaus verändert.

Tabelle 2.3.5.

*Ergebnisse Regressionsanalysen mit Bildungsniveau für Frauen und Männer – Bedingung GV<sup>1</sup>*

Merkmal und Item-Nr.	Frauen			Männer		
	$\beta$	$z$	$p$	$\beta$	$z$	$p$
Alter – 1						
Reale Werte	0.071	3.97	<b>&lt; .001</b>	0.058	2.13	<b>.03</b>
Kein Abitur	-0.105	-0.46	.64	-0.588	-1.33	.18
Hochschulabschluss	-0.061	-0.24	.80	-0.317	-0.70	.47
Reale Werte : Kein Abitur	-0.006	-0.32	.74	0.023	0.72	.47
Reale Werte : Hochschulabschluss	-0.022	-0.87	.38	0.029	0.80	.42
Alter – 1(a)						
Reale Werte	0.055	3.25	<b>.001</b>	0.030	1.11	.26
Kein Abitur	-0.344	-1.55	.11	-0.658	-1.49	.13
Hochschulabschluss	0.034	0.14	.88	-0.281	-0.62	.53
Reale Werte : Kein Abitur	-0.041	-2.03	<b>.04</b>	0.007	0.25	.80
Reale Werte : Hochschulabschluss	-0.048	-1.93	<b>.05</b>	0.007	0.19	.84
Alter – 1(b)						
Reale Werte	0.056	3.22	<b>.001</b>	0.033	1.28	.19
Kein Abitur	-0.205	-0.91	.35	-0.088	-0.20	.83
Hochschulabschluss	0.034	0.13	.89	-0.256	-0.54	.58
Reale Werte : Kein Abitur	-0.025	-1.22	.22	0.007	0.24	.80
Reale Werte : Hochschulabschluss	-0.051	-2.04	<b>.04</b>	0.027	0.74	.45
Körpergröße – 2						
Reale Werte	0.420	11.35	<b>&lt; .001</b>	0.419	6.04	<b>&lt; .001</b>
Kein Abitur	0.564	2.40	<b>.01</b>	0.044	0.09	.92
Hochschulabschluss	0.279	1.07	.28	-0.464	-0.94	.34
Reale Werte : Kein Abitur	-0.068	-1.66	.09	-0.165	-2.16	<b>.03</b>
Reale Werte : Hochschulabschluss	-0.052	-1.18	.23	-0.053	-0.67	.50
Körpergröße – 2(a)						
Reale Werte	0.312	9.60	<b>&lt; .001</b>	0.313	5.79	<b>&lt; .001</b>
Kein Abitur	0.217	0.98	.32	-0.123	-0.27	.78
Hochschulabschluss	0.081	0.32	.74	-0.163	-0.34	.73
Reale Werte : Kein Abitur	-0.072	-1.87	.06	-0.074	-1.16	.24
Reale Werte : Hochschulabschluss	-0.003	-0.08	.93	-0.005	-0.07	.94
Körpergröße – 2(b)						
Reale Werte	0.439	11.16	<b>&lt; .001</b>	0.438	6.23	<b>&lt; .001</b>
Kein Abitur	0.421	1.82	.06	0.099	0.20	.83
Hochschulabschluss	0.006	0.02	.97	-0.538	-1.07	.28
Reale Werte : Kein Abitur	-0.132	-3.07	<b>.002</b>	-0.068	-0.92	.35
Reale Werte : Hochschulabschluss	-0.078	-1.65	.09	-0.083	-1.06	.28

Tabelle 2.3.5. (Fortsetzung)

*Ergebnisse Regressionsanalysen mit Bildungsniveau für Frauen und Männer – Bedingung GV<sup>1</sup>*

Merkmal und Item-Nr.	Frauen			Männer		
	$\beta$	$z$	$p$	$\beta$	$z$	$p$
Körpergewicht (BMI) – 4						
Reale Werte	0.444	9.89	<b>&lt; .001</b>	0.519	5.35	<b>&lt; .001</b>
Kein Abitur	-0.080	-0.34	.73	0.364	0.83	.40
Hochschulabschluss	0.317	1.10	.26	0.484	1.04	.29
Reale Werte : Kein Abitur	-0.063	-1.30	.19	-0.081	-0.71	.47
Reale Werte : Hochschulabschluss	-0.060	-0.91	.36	-0.243	-2.07	<b>.03</b>
Körpergewicht (BMI) – 4(a)						
Reale Werte	0.317	7.57	<b>&lt; .001</b>	0.618	5.14	<b>&lt; .001</b>
Kein Abitur	-0.262	-1.13	.25	0.362	0.79	.42
Hochschulabschluss	0.244	0.82	.40	0.609	1.29	.19
Reale Werte : Kein Abitur	-0.039	-0.83	.40	-0.181	-1.33	.18
Reale Werte : Hochschulabschluss	0.048	0.70	.48	-0.393	-2.96	<b>.003</b>
Körpergewicht (BMI) – 4(b)						
Reale Werte	0.404	9.95	<b>&lt; .001</b>	0.648	5.98	<b>&lt; .001</b>
Kein Abitur	-0.386	-1.69	.09	0.146	0.32	.74
Hochschulabschluss	0.029	0.11	.91	0.461	0.96	.33
Reale Werte : Kein Abitur	-0.051	-1.17	.24	-0.136	-1.14	.25
Reale Werte : Hochschulabschluss	-0.048	-0.83	.40	-0.203	-1.60	.10
Einkommen – 5						
Reale Werte	1.253	5.25	<b>&lt; .001</b>	1.612	4.34	<b>&lt; .001</b>
Kein Abitur	-0.267	-1.18	.23	0.061	0.12	.89
Hochschulabschluss	-0.118	-0.47	.63	-0.341	-0.66	.50
Reale Werte : Kein Abitur	0.347	1.10	.27	0.178	0.39	.69
Reale Werte : Hochschulabschluss	-0.156	-0.54	.58	0.318	0.69	.48
Einkommen – 5(a)						
Reale Werte	0.871	3.95	<b>&lt; .001</b>	1.322	3.80	<b>&lt; .001</b>
Kein Abitur	0.029	0.13	.89	0.224	0.48	.63
Hochschulabschluss	0.159	0.63	.52	-0.225	-0.46	.64
Reale Werte : Kein Abitur	0.332	1.12	.25	0.258	0.60	.54
Reale Werte : Hochschulabschluss	0.094	0.33	.73	-0.224	-0.57	.56
Einkommen – 5(b)						
Reale Werte	1.253	5.22	<b>&lt; .001</b>	1.574	4.66	<b>&lt; .001</b>
Kein Abitur	0.027	0.12	.90	-0.542	-1.20	.22
Hochschulabschluss	0.048	0.19	.84	-1.054	-2.20	<b>.02</b>
Reale Werte : Kein Abitur	0.361	1.14	.25	-0.685	-1.66	.09
Reale Werte : Hochschulabschluss	-0.372	-1.30	.19	-0.856	-2.40	<b>.01</b>

**Anmerkungen.** $\beta$  = Regressionsgewicht; Reale Werte = Effekt für den Prädiktor *reale Werte* in der Referenzgruppe *Abitur*.<sup>1</sup> Anzeige Balkendiagramm mit der Gesamtmerkmalsverteilung in der Population.

(a) Egozentrische Perspektive; (b) Allozentrische Perspektive.

Signifikante Ergebnisse ( $p \leq .05$ ) sind fett gedruckt.

Tabelle 2.3.6.

*Ergebnisse Regressionsanalysen mit Bildungsniveau für Frauen und Männer – Bedingung GSV<sup>1</sup>*

Merkmal und Item-Nr.	Frauen			Männer		
	$\beta$	$z$	$p$	$\beta$	$z$	$p$
Alter – 1						
Reale Werte	0.078	3.63	<b>&lt; .001</b>	-0.034	-0.84	.39
Kein Abitur	0.430	1.73	.08	1.158	2.20	<b>.02</b>
Hochschulabschluss	0.506	1.95	<b>.05</b>	0.884	1.62	.10
Reale Werte : Kein Abitur	-0.042	-1.70	.08	0.084	1.84	.06
Reale Werte : Hochschulabschluss	0.006	0.23	.81	0.033	0.66	.50
Alter – 1(a)						
Reale Werte	0.022	1.10	.26	-0.047	-1.17	.23
Kein Abitur	0.311	1.28	.19	0.887	1.66	.09
Hochschulabschluss	0.587	2.32	<b>.02</b>	0.786	1.43	.15
Reale Werte : Kein Abitur	-0.008	-0.35	.72	0.045	1.00	.31
Reale Werte : Hochschulabschluss	-0.010	-0.37	.70	0.089	1.81	.06
Alter – 1(b)						
Reale Werte	0.041	1.93	<b>.05</b>	0.005	0.17	.86
Kein Abitur	0.098	0.39	.69	0.766	1.57	.11
Hochschulabschluss	0.141	0.54	.58	0.547	1.11	.26
Reale Werte : Kein Abitur	0.005	0.22	.81	0.032	0.81	.41
Reale Werte : Hochschulabschluss	-0.003	-0.11	.90	0.009	0.21	.83
Körpergröße – 2						
Reale Werte	0.376	10.75	<b>&lt; .001</b>	0.241	4.18	<b>&lt; .001</b>
Kein Abitur	-0.214	-0.90	.36	0.134	0.28	.77
Hochschulabschluss	-0.457	-1.80	.07	0.224	0.44	.65
Reale Werte : Kein Abitur	-0.054	-1.39	.16	-0.043	-0.65	.51
Reale Werte : Hochschulabschluss	0.121	2.63	<b>.008</b>	0.049	0.71	.47
Körpergröße – 2(a)						
Reale Werte	0.351	10.23	<b>&lt; .001</b>	0.142	2.70	<b>.006</b>
Kein Abitur	0.215	0.92	.35	0.685	1.52	.12
Hochschulabschluss	-0.361	-1.43	.15	0.768	1.56	.11
Reale Werte : Kein Abitur	-0.116	-2.95	<b>.003</b>	-0.019	-0.29	.76
Reale Werte : Hochschulabschluss	0.068	1.48	.13	0.119	1.74	.08
Körpergröße – 2(b)						
Reale Werte	0.350	10.45	<b>&lt; .001</b>	0.275	4.75	<b>&lt; .001</b>
Kein Abitur	-0.105	-0.45	.65	0.282	0.60	.54
Hochschulabschluss	-0.514	-2.07	<b>.03</b>	0.172	0.34	.72
Reale Werte : Kein Abitur	-0.028	-0.71	.47	-0.097	-1.45	.14
Reale Werte : Hochschulabschluss	0.071	1.59	.11	0.035	0.50	.61



Tabelle 2.3.6. (Fortsetzung)

*Ergebnisse Regressionsanalysen mit Bildungsniveau für Frauen und Männer – Bedingung GSV<sup>1</sup>*

Merkmal und Item-Nr.	Frauen			Männer		
	$\beta$	$z$	$p$	$\beta$	$z$	$p$
Körpergewicht – 4						
Reale Werte	0.486	8.70	<b>&lt; .001</b>	0.366	3.94	<b>&lt; .001</b>
Kein Abitur	-0.123	-0.50	.61	0.332	0.69	.48
Hochschulabschluss	-0.152	-0.58	.56	0.692	1.32	.18
Reale Werte : Kein Abitur	-0.024	-0.39	.69	-0.002	-0.02	.98
Reale Werte : Hochschulabschluss	-0.056	-0.85	.39	0.148	1.14	.25
Körpergewicht – 4(a)						
Reale Werte	0.395	7.53	<b>&lt; .001</b>	0.255	3.31	<b>&lt; .001</b>
Kein Abitur	-0.101	-0.42	.67	0.591	1.28	.19
Hochschulabschluss	0.106	0.39	.69	1.182	2.32	<b>.02</b>
Reale Werte : Kein Abitur	-0.016	-0.28	.77	0.079	0.85	.39
Reale Werte : Hochschulabschluss	0.011	0.17	.85	0.185	1.49	.13
Körpergewicht – 4(b)						
Reale Werte	0.391	8.15	<b>&lt; .001</b>	0.350	4.05	<b>&lt; .001</b>
Kein Abitur	-0.215	-0.91	.35	0.501	1.07	.28
Hochschulabschluss	0.400	1.56	.11	0.687	1.35	.17
Reale Werte : Kein Abitur	-0.022	-0.42	.67	0.098	0.97	.33
Reale Werte : Hochschulabschluss	0.036	0.60	.54	0.186	1.49	.13
Einkommen – 5						
Reale Werte	1.389	5.98	<b>&lt; .001</b>	1.340	3.24	<b>.001</b>
Kein Abitur	0.046	0.19	.84	-0.183	-0.36	.71
Hochschulabschluss	-0.445	-1.75	.07	-0.309	-0.57	.56
Reale Werte : Kein Abitur	0.426	1.31	.18	-0.333	-0.67	.50
Reale Werte : Hochschulabschluss	-0.124	-0.45	.65	-0.000	-0.00	.99
Einkommen – 5(a)						
Reale Werte	1.111	4.97	<b>&lt; .001</b>	0.537	1.21	.22
Kein Abitur	-0.132	-0.57	.56	-0.103	-0.19	.84
Hochschulabschluss	-0.639	-2.51	<b>.01</b>	-0.068	-0.12	.90
Reale Werte : Kein Abitur	0.363	1.17	.24	0.288	0.53	.59
Reale Werte : Hochschulabschluss	-0.175	-0.64	.51	0.321	0.65	.51
Einkommen – 5(b)						
Reale Werte	1.019	4.40	<b>&lt; .001</b>	1.524	3.43	<b>&lt; .001</b>
Kein Abitur	-0.009	-0.04	.96	-0.339	-0.67	.50
Hochschulabschluss	-0.298	-1.19	.23	-0.377	-0.70	.48
Reale Werte : Kein Abitur	0.840	2.51	<b>.01</b>	-0.037	-0.07	.94
Reale Werte : Hochschulabschluss	-0.043	-0.15	.87	-0.294	-0.60	.54

*Anmerkungen.* $\beta$  = Regressionsgewicht; Reale Werte = Effekt für den Prädiktor *reale Werte* in der Referenzgruppe *Abitur*.<sup>1</sup> Anzeige Balkendiagramm mit der geschlechtsspezifischen Merkmalsverteilung in der Population.

(a) Egozentrische Perspektive; (b) Allozentrische Perspektive.

Signifikante Ergebnisse ( $p \leq .05$ ) sind fett gedruckt.

Tabelle 2.3.7.

*Ergebnisse Regressionsanalysen mit Bildungsniveau für Frauen und Männer – Bedingung GV/GSV<sup>1</sup>*

Merkmal und Item-Nr.	Frauen			Männer		
	$\beta$	$z$	$p$	$\beta$	$z$	$p$
Alter – 1						
Reale Werte	0.063	3.10	<b>.001</b>	0.063	2.12	<b>.03</b>
Kein Abitur	0.045	0.18	.85	0.124	0.29	.77
Hochschulabschluss	-0.091	-0.36	.71	0.247	0.54	.58
Reale Werte : Kein Abitur	0.011	0.46	.64	-0.012	-0.33	.73
Reale Werte : Hochschulabschluss	-0.002	-0.09	.92	0.032	0.86	.38
Alter – 1(a)						
Reale Werte	0.023	1.14	.25	0.018	0.61	.53
Kein Abitur	0.081	0.33	.73	0.096	0.23	.81
Hochschulabschluss	0.058	0.23	.81	0.227	0.50	.61
Reale Werte : Kein Abitur	0.005	0.23	.81	-0.022	-0.59	.55
Reale Werte : Hochschulabschluss	-0.000	-0.00	.99	0.037	1.02	.30
Alter – 1(b)						
Reale Werte	0.065	3.33	<b>&lt; .001</b>	0.043	1.48	.13
Kein Abitur	0.072	0.30	.76	-0.037	-0.09	.92
Hochschulabschluss	0.034	0.13	.89	0.488	1.07	.28
Reale Werte : Kein Abitur	-0.025	-1.12	.25	0.019	0.55	.58
Reale Werte : Hochschulabschluss	-0.005	-0.19	.84	0.035	0.96	.33
Körpergröße – 2						
Reale Werte	0.436	10.67	<b>&lt; .001</b>	0.300	5.16	<b>&lt; .001</b>
Kein Abitur	0.085	0.35	.72	0.141	0.30	.76
Hochschulabschluss	0.166	0.64	.51	0.458	0.93	.35
Reale Werte : Kein Abitur	-0.029	-0.66	.50	0.029	0.41	.68
Reale Werte : Hochschulabschluss	-0.072	-1.53	.12	0.074	1.03	.30
Körpergröße – 2(a)						
Reale Werte	0.409	10.04	<b>&lt; .001</b>	0.240	4.76	<b>&lt; .001</b>
Kein Abitur	-0.284	-1.17	.23	0.228	0.51	.61
Hochschulabschluss	-0.013	-0.05	.95	0.305	0.65	.51
Reale Werte : Kein Abitur	-0.011	-0.25	.79	0.029	0.43	.66
Reale Werte : Hochschulabschluss	-0.097	-2.05	<b>.04</b>	0.099	1.46	.14
Körpergröße – 2(b)						
Reale Werte	0.470	10.30	<b>&lt; .001</b>	0.337	5.68	<b>&lt; .001</b>
Kein Abitur	0.156	0.62	.53	0.269	0.57	.56
Hochschulabschluss	0.392	1.47	.14	0.579	1.18	.23
Reale Werte : Kein Abitur	-0.053	-1.07	.28	0.017	0.23	.81
Reale Werte : Hochschulabschluss	-0.092	-1.78	.07	0.085	1.14	.25

Tabelle 2.3.7. (Fortsetzung)

*Ergebnisse Regressionsanalysen mit Bildungsniveau für Frauen und Männer – Bedingung GV/GSV<sup>1</sup>*

Merkmal und Item-Nr.	Frauen			Männer		
	$\beta$	$z$	$p$	$\beta$	$z$	$p$
Körpergewicht – 4						
Reale Werte	0.470	8.75	<b>&lt; .001</b>	0.516	5.34	<b>&lt; .001</b>
Kein Abitur	-0.262	-1.05	.29	0.015	0.03	.97
Hochschulabschluss	-0.397	-1.38	.16	-0.590	-1.18	.23
Reale Werte : Kein Abitur	-0.061	-1.07	.28	-0.026	-0.24	.80
Reale Werte : Hochschulabschluss	-0.030	-0.46	.64	-0.191	-1.49	.13
Körpergewicht – 4(a)						
Reale Werte	0.377	7.15	<b>&lt; .001</b>	0.384	4.44	<b>&lt; .001</b>
Kein Abitur	-0.316	-1.24	.21	0.526	1.22	.22
Hochschulabschluss	-0.103	-0.34	.73	-0.556	-1.15	.24
Reale Werte : Kein Abitur	0.011	0.19	.84	0.026	0.24	.80
Reale Werte : Hochschulabschluss	0.041	0.59	.55	-0.158	-1.30	.19
Körpergewicht – 4(b)						
Reale Werte	0.400	8.14	<b>&lt; .001</b>	0.485	5.47	<b>&lt; .001</b>
Kein Abitur	-0.117	-0.48	.62	-0.006	-0.01	.98
Hochschulabschluss	0.131	0.47	.63	-0.678	-1.33	.18
Reale Werte : Kein Abitur	-0.022	-0.42	.66	-0.133	-1.30	.19
Reale Werte : Hochschulabschluss	0.031	0.51	.60	-0.295	-2.23	<b>.02</b>
Einkommen – 5						
Reale Werte	1.538	6.68	<b>&lt; .001</b>	0.969	3.49	<b>&lt; .001</b>
Kein Abitur	0.354	1.51	.13	-0.867	-1.97	<b>.04</b>
Hochschulabschluss	-0.036	-0.14	.88	-0.592	-1.20	.22
Reale Werte : Kein Abitur	-0.002	-0.00	.99	0.461	1.22	.21
Reale Werte : Hochschulabschluss	-0.499	-1.88	<b>.05</b>	0.477	1.30	.19
Einkommen – 5(a)						
Reale Werte	1.107	5.29	<b>&lt; .001</b>	0.854	3.27	<b>.001</b>
Kein Abitur	0.377	1.59	.11	-0.586	-1.35	.17
Hochschulabschluss	0.260	1.03	.30	-0.434	-0.92	.35
Reale Werte : Kein Abitur	0.133	0.47	.63	0.488	1.41	.15
Reale Werte : Hochschulabschluss	-0.313	-1.27	.20	0.538	1.57	.11
Einkommen – 5(b)						
Reale Werte	1.440	6.01	<b>&lt; .001</b>	1.175	4.20	<b>&lt; .001</b>
Kein Abitur	0.215	0.88	.37	-0.384	-0.89	.36
Hochschulabschluss	-0.062	-0.24	.81	-0.887	-1.85	.06
Reale Werte : Kein Abitur	-0.074	-0.24	.80	0.296	0.84	.40
Reale Werte : Hochschulabschluss	-0.658	-2.46	<b>.01</b>	0.239	0.69	.48

*Anmerkungen.* $\beta$  = Regressionsgewicht; Reale Werte = Effekt für den Prädiktor *reale Werte* in der Referenzgruppe *Abitur*.<sup>1</sup> Gleichzeitige Anzeige Balkendiagramme mit der Gesamtmerkmalsverteilung und der geschlechts-spezifischen Merkmalsverteilung in der Population; (a) Egozentrische Perspektive; (b) Allozentrische Perspektive; signifikante Ergebnisse ( $p \leq .05$ ) sind fett gedruckt.

Tabelle 2.3.8.

*Ergebnisse Regressionsanalysen mit Bildungsniveau für Frauen und Männer – Bedingung KV<sup>1</sup>*

Merkmal und Item-Nr.	Frauen			Männer		
	$\beta$	$z$	$p$	$\beta$	$z$	$p$
Alter – 1						
Reale Werte	0.064	3.66	<b>&lt; .001</b>	0.061	2.03	<b>.04</b>
Kein Abitur	0.111	0.49	.62	1.207	2.58	<b>.009</b>
Hochschulabschluss	0.112	0.46	.64	1.233	2.55	<b>.01</b>
Reale Werte : Kein Abitur	0.002	0.11	.91	0.012	0.33	.74
Reale Werte : Hochschulabschluss	0.014	0.59	.55	0.028	0.75	.44
Alter – 1(a)						
Reale Werte	0.022	1.28	.19	-0.012	-0.43	.66
Kein Abitur	0.199	0.89	.37	0.384	0.88	.37
Hochschulabschluss	0.087	0.36	.71	0.348	0.77	.43
Reale Werte : Kein Abitur	-0.009	-0.43	.66	0.047	1.40	.16
Reale Werte : Hochschulabschluss	0.009	0.37	.70	0.054	1.54	.12
Alter – 1(b)						
Reale Werte	0.050	2.73	<b>.006</b>	0.024	0.87	.38
Kein Abitur	0.018	0.08	.93	0.695	1.57	.11
Hochschulabschluss	-0.314	-1.26	.20	0.747	1.68	.09
Reale Werte : Kein Abitur	-0.031	-1.38	.16	0.033	0.95	.33
Reale Werte : Hochschulabschluss	0.012	0.49	.61	0.022	0.64	.51
Körpergröße – 2						
Reale Werte	0.420	12.14	<b>&lt; .001</b>	0.402	5.54	<b>&lt; .001</b>
Kein Abitur	0.301	1.28	.19	-0.076	-0.15	.87
Hochschulabschluss	0.015	0.05	.95	-0.175	-0.35	.72
Reale Werte : Kein Abitur	-0.014	-0.36	.71	-0.077	-0.93	.35
Reale Werte : Hochschulabschluss	-0.016	-0.38	.70	0.084	0.88	.37
Körpergröße – 2(a)						
Reale Werte	0.333	10.69	<b>&lt; .001</b>	0.183	3.01	<b>.002</b>
Kein Abitur	0.008	0.03	.96	-0.347	-0.76	.44
Hochschulabschluss	0.005	0.02	.98	-0.359	-0.78	.43
Reale Werte : Kein Abitur	-0.060	-1.56	.11	0.076	0.98	.32
Reale Werte : Hochschulabschluss	-0.014	-0.34	.73	0.182	2.09	<b>.03</b>
Körpergröße – 2(b)						
Reale Werte	0.387	11.70	<b>&lt; .001</b>	0.277	4.45	<b>&lt; .001</b>
Kein Abitur	0.041	0.17	.85	0.636	1.37	.17
Hochschulabschluss	0.123	0.48	.63	-0.033	-0.07	.94
Reale Werte : Kein Abitur	-0.016	-0.41	.68	0.039	0.51	.60
Reale Werte : Hochschulabschluss	-0.012	-0.27	.78	0.105	1.23	.21

Tabelle 2.3.8. (Fortsetzung)

*Ergebnisse Regressionsanalysen mit Bildungsniveau für Frauen und Männer – Bedingung KV<sup>1</sup>*

Merkmal und Item-Nr.	Frauen			Männer		
	$\beta$	$z$	$p$	$\beta$	$z$	$p$
Körpergewicht – 4						
Reale Werte	0.389	8.98	<b>&lt; .001</b>	0.444	5.35	<b>&lt; .001</b>
Kein Abitur	-0.058	-0.24	.80	0.104	0.23	.81
Hochschulabschluss	0.205	0.70	.47	0.448	0.96	.33
Reale Werte : Kein Abitur	0.099	1.88	.06	-0.046	-0.49	.62
Reale Werte : Hochschulabschluss	0.131	1.95	<b>.05</b>	0.104	0.94	.34
Körpergewicht – 4(a)						
Reale Werte	0.287	6.97	<b>&lt; .001</b>	0.395	5.19	<b>&lt; .001</b>
Kein Abitur	-0.111	-0.46	.64	0.848	1.86	.06
Hochschulabschluss	-0.123	-0.44	.66	0.957	2.04	<b>.04</b>
Reale Werte : Kein Abitur	0.017	0.34	.73	-0.055	-0.57	.56
Reale Werte : Hochschulabschluss	0.041	0.66	.50	0.117	1.05	.29
Körpergewicht – 4(b)						
Reale Werte	0.319	8.41	<b>&lt; .001</b>	0.399	5.10	<b>&lt; .001</b>
Kein Abitur	-0.000	-0.00	.99	-0.267	-0.58	.55
Hochschulabschluss	0.600	2.22	<b>.02</b>	0.302	0.64	.51
Reale Werte : Kein Abitur	0.009	0.20	.83	-0.008	-0.09	.92
Reale Werte : Hochschulabschluss	0.160	2.74	<b>.006</b>	0.071	0.62	.53
Einkommen – 5						
Reale Werte	1.340	7.37	<b>&lt; .001</b>	0.837	3.81	<b>&lt; .001</b>
Kein Abitur	0.037	0.16	.86	0.235	0.52	.59
Hochschulabschluss	0.184	0.75	.45	0.322	0.69	.48
Reale Werte : Kein Abitur	0.138	0.52	.59	0.590	1.49	.13
Reale Werte : Hochschulabschluss	-0.376	-1.52	.12	-0.507	-2.22	<b>.02</b>
Einkommen – 5(a)						
Reale Werte	1.215	6.86	<b>&lt; .001</b>	1.089	4.34	<b>&lt; .001</b>
Kein Abitur	0.078	0.35	.72	-0.320	-0.69	.48
Hochschulabschluss	0.159	0.66	.50	-0.408	-0.89	.37
Reale Werte : Kein Abitur	0.067	0.26	.79	-0.025	-0.06	.95
Reale Werte : Hochschulabschluss	-0.470	-2.00	<b>.04</b>	-0.757	-2.98	<b>.002</b>
Einkommen – 5(b)						
Reale Werte	1.200	6.60	<b>&lt; .001</b>	0.752	3.47	<b>&lt; .001</b>
Kein Abitur	0.230	1.04	.29	0.814	1.75	.07
Hochschulabschluss	0.219	0.91	.35	0.193	0.43	.66
Reale Werte : Kein Abitur	0.375	1.37	.17	0.705	1.77	.07
Reale Werte : Hochschulabschluss	-0.261	-1.05	.29	-0.585	-2.52	<b>.01</b>

*Anmerkungen.* $\beta$  = Regressionsgewicht; Reale Werte = Effekt für den Prädiktor *reale Werte* in der Referenzgruppe *Abitur*.<sup>1</sup> Anzeige keiner Merkmalsverteilung. Bei der Bedingung KV handelt es sich um die Kontrollbedingung.

(a) Egozentrische Perspektive; (b) Allozentrische Perspektive.

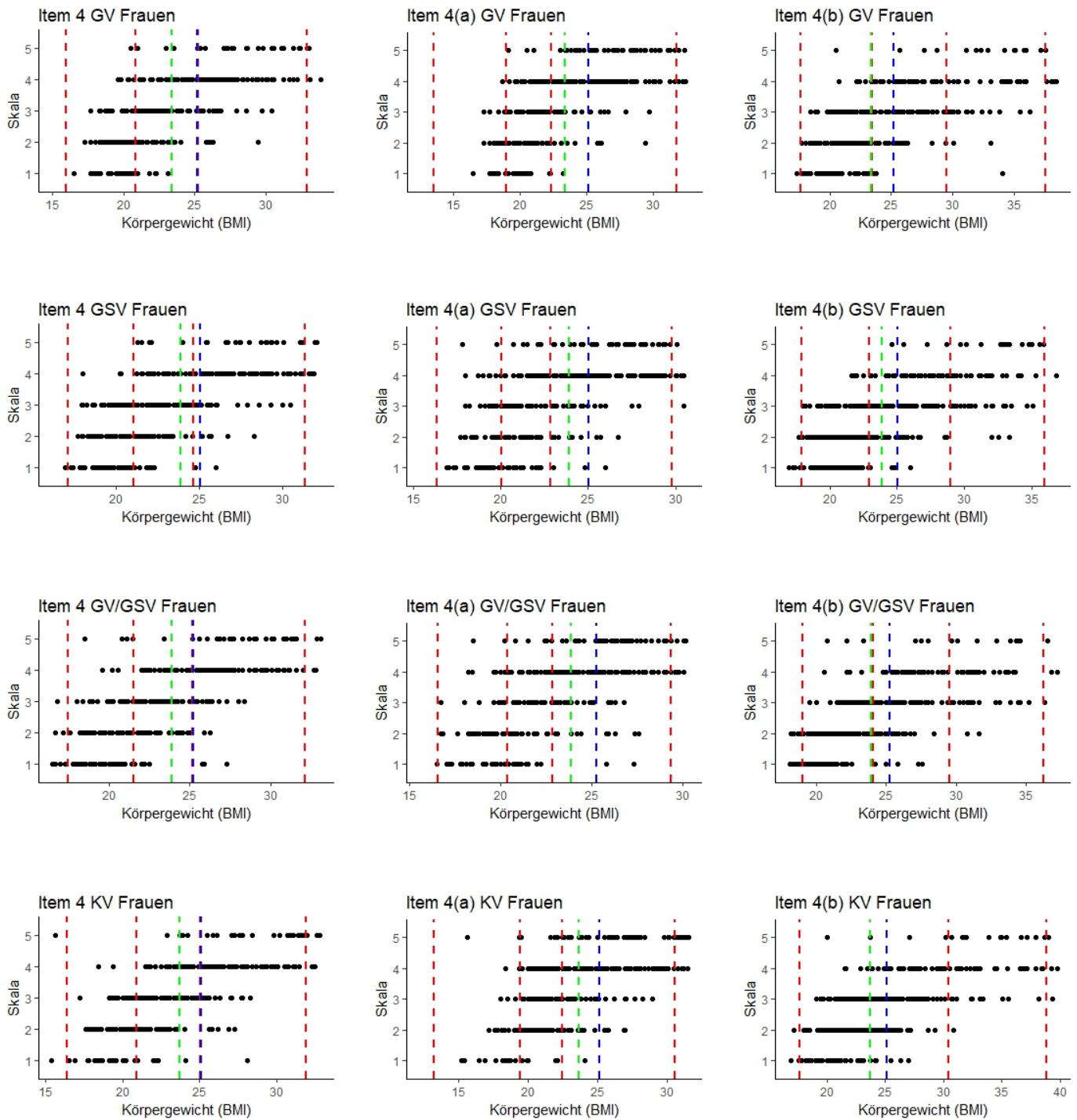
Signifikante Ergebnisse ( $p \leq .05$ ) sind fett gedruckt.

*Körpergewicht.* In den Abbildungen 2.3.1. und 2.3.2. ist für das Merkmal Körpergewicht für jede Bedingung getrennt nach Geschlecht die Lage der aus den Proportional Odds Regressionen ermittelten Schwellenparameter graphisch (rot) veranschaulicht. Mittelwert (blau) und Median (grün) der realen Werte sind für die jeweilige Teilstichprobe des nicht-messwiederholten Faktors abgebildet. In Fällen, in denen der Mittelwert oder der Median auf einer Schwelle liegen, konnte aus technischen Gründen nur der Mittelwert bzw. der Median angezeigt werden. Die Tabellen 2.3.9. bis 2.3.12. berichten die Schwellenparameter und Steigungskoeffizienten (flexibles Modell) für jede Bedingung getrennt nach Geschlecht. Die Ergebnisse der Likelihood Ratio Tests sind in Anhang A einzusehen (Tabellen A.2. und A.3.). Die Ergebnisse werden für Körpergewicht in Kilogramm (entnommen aus Eichhorn, 2016) und den BMI berichtet.

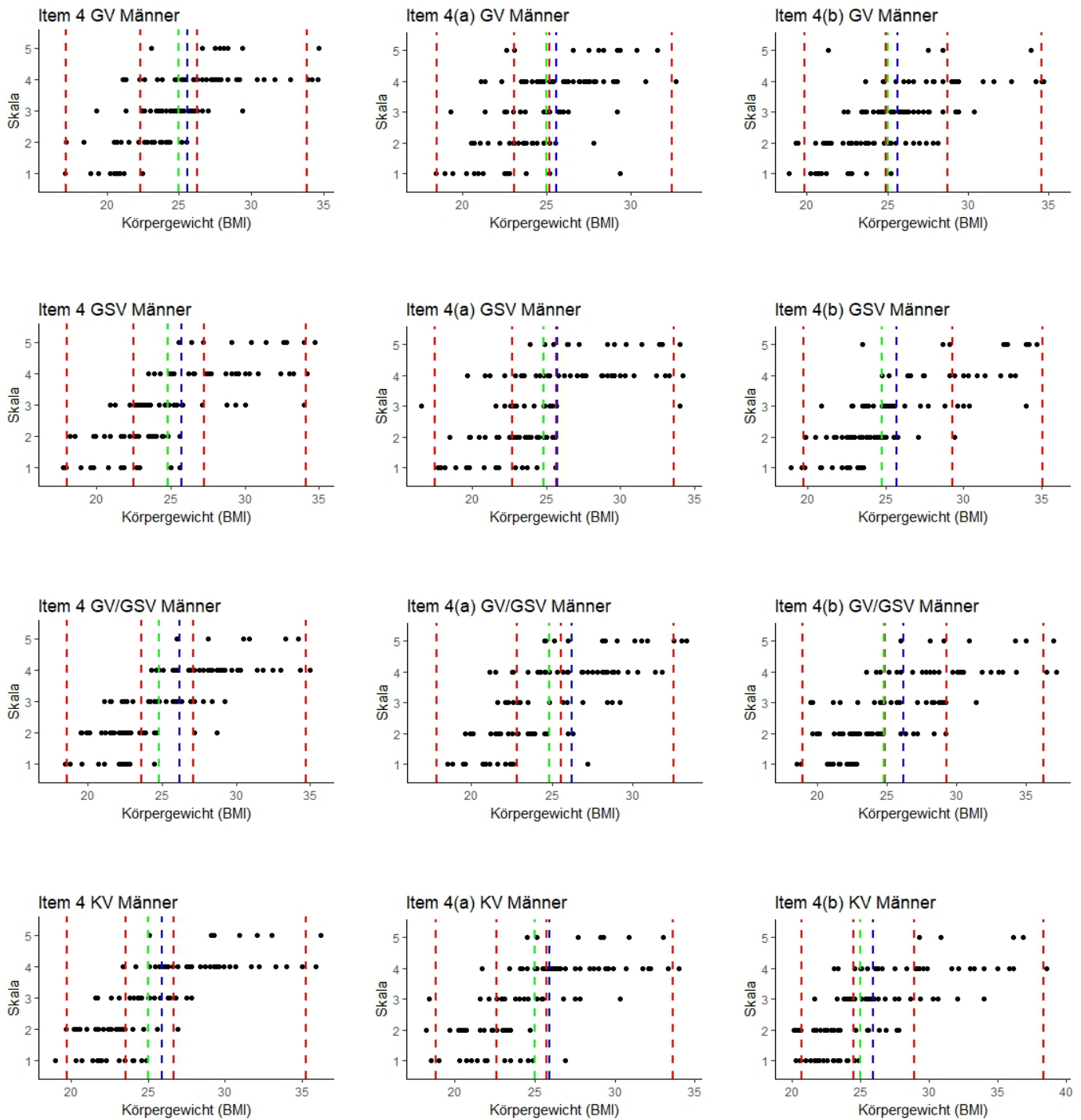
Abbildung 2.3.1. zeigt, dass bei der weiblichen Teilstichprobe Mittelwert und Median bei egozentrischer Formulierung innerhalb der vierten Antwortkategorie liegen und bei allozentrischer Formulierung in der mittleren Antwortkategorie liegen mit einer Tendenz zur unteren Kategoriengrenze. Bei der Formulierung ohne implizite Perspektive liegen Mittelwert und Median innerhalb der mittleren Antwortkategorie mit einer Tendenz zur oberen Kategoriengrenze. Zwischen den Bedingungen GV, GSV, GV/GSV und KV sind keine Unterschiede sichtbar. Es ergibt sich unter keiner Bedingung eine Annäherung an äquidistante Schwellen (elf von zwölf Likelihood Ratio Tests sind signifikant).

Abbildung 2.3.2. zeigt, dass in der männlichen Teilstichprobe Mittelwert und Median bei den Formulierungen ohne implizite Perspektive sowie mit allozentrischer Perspektive innerhalb der mittleren Antwortkategorie liegen. Bei der egozentrischen Formulierung liegen Mittelwert und Median nahe der dritten Schwelle, entweder in der dritten oder vierten Antwortkategorie. Nur bei der Formulierung mit allozentrischer Perspektive in Verbindung mit den Bedingungen GV, GSV und GV/GSV ergeben sich annähernd äquidistante Schwellen. Die entsprechenden Likelihood Ratio Tests sind nicht signifikant.

Bei Frauen ergeben die BMI-basierten Regressionsanalysen im Vergleich zu denen auf Basis des absoluten Körpergewichts nur eine Änderung in den Schwellenabständen. So entfällt bei BMI-basierten Regressionsanalysen mit allozentrischer Formulierung die Annäherung an äquidistante Schwellen. Bei der männlichen Teilstichprobe fallen die Ergebnisse beider Regressionsanalysen vergleichbar aus.



**Abbildung 2.3.1.** Regression Antwortskala und reale Werte für das Merkmal Körpergewicht für Frauen. Körpergewicht angegeben als BMI; schwarze Punkte = gewählte Kategorien; rote Linien = geschätzte Schwellen zwischen den Kategorien der Skala; blaue Linie = Mittelwert der realen Werte; grüne Linie = Median der realen Werte. GV = Anzeige Balkendiagramm mit der Gesamtmerkmalsverteilung in der Population. GSV = Anzeige Balkendiagramm mit der geschlechtsspezifischen Merkmalsverteilung in der Population. GV/GSV = Gleichzeitige Anzeige Balkendiagramme mit der Gesamtmerkmalsverteilung und der geschlechtsspezifischen Merkmalsverteilung in der Population. KV = Anzeige keiner Merkmalsverteilung. Bei der Bedingung KV handelt es sich um die Kontrollbedingung. (a) Egozentrische Perspektive; (b) Allozentrische Perspektive.



**Abbildung 2.3.2.** Regression Antwortskala und reale Werte für das Merkmal Körpergewicht für Männer. Körpergewicht angegeben als BMI; schwarze Punkte = gewählte Kategorien; rote Linien = geschätzte Schwellen zwischen den Kategorien der Skala; blaue Linie = Mittelwert der realen Werte; grüne Linie = Median der realen Werte. GV = Anzeige Balkendiagramm mit der Gesamtmerkmalsverteilung in der Population. GSV = Anzeige Balkendiagramm mit der geschlechtsspezifischen Merkmalsverteilung in der Population. GV/GSV = Gleichzeitige Anzeige Balkendiagramme mit der Gesamtmerkmalsverteilung und der geschlechtsspezifischen Merkmalsverteilung in der Population. KV = Anzeige keiner Merkmalsverteilung. Bei der Bedingung KV handelt es sich um die Kontrollbedingung. (a) Egozentrische Perspektive; (b) Allozentrische Perspektive.



Tabelle 2.3.9.

*Schwellenparameter ( $t^1$ ) und Steigung ( $s^2$ ) für Körpergewicht der Bedingung GV<sup>3</sup> für Frauen und Männer*

Merkmal und Item-Nr.	Frauen					Männer				
	$t_1$	$t_2$	$t_3$	$t_4$	$s$	$t_1$	$t_2$	$t_3$	$t_4$	$s$
Körpergewicht (kg) <sup>4</sup>										
4	5.769	7.712	9.431	12.378	0.133	6.211	8.334	9.929	12.848	0.115
4(a)	3.463	5.054	6.005	8.693	0.097	6.094	7.915	8.710	11.418	0.106
4(b)	5.448	7.521	9.609	12.337	0.115	9.085	11.738	13.733	16.343	0.145
Körpergewicht (BMI)										
4	6.255	8.169	9.881	12.871	0.391	7.034	9.143	10.759	13.844	0.409
4(a)	3.962	5.552	6.526	9.276	0.292	7.236	9.059	9.878	12.761	0.392
4(b)	6.248	8.328	10.491	13.342	0.355	10.514	13.197	15.228	18.320	0.529

*Anmerkungen.*<sup>1</sup> Schwellenparameter = Punkt, an dem die Wahrscheinlichkeit für die Wahl der nächsthöheren Antwortkategorie größer wird als für die Wahl der vorherigen Antwortkategorie.<sup>2</sup> Steigung = Steigungskoeffizient der Regression.<sup>3</sup> Anzeige Balkendiagramm mit der Gesamtmerkmalsverteilung in der Population.<sup>4</sup> Die Ergebnisse für Körpergewicht in Kilogramm wurden aus Eichhorn (2016) entnommen. $t_1$  = erste Schwelle;  $t_2$  = zweite Schwelle;  $t_3$  = dritte Schwelle;  $t_4$  = vierte Schwelle.

Tabelle 2.3.10.

*Schwellenparameter ( $t^1$ ) und Steigung ( $s^2$ ) für Körpergewicht der Bedingung GSV<sup>3</sup> für Frauen und Männer*

Merkmal und Item-Nr.	Frauen					Männer				
	$t_1$	$t_2$	$t_3$	$t_4$	$s$	$t_1$	$t_2$	$t_3$	$t_4$	$s$
Körpergewicht (kg) <sup>4</sup>										
4	6.275	8.033	9.603	12.410	0.137	7.180	9.070	10.996	13.748	0.122
4(a)	5.465	6.895	7.982	10.544	0.124	5.488	7.229	8.176	10.756	0.096
4(b)	5.817	7.679	9.876	12.333	0.119	6.839	8.954	10.715	12.923	0.109
Körpergewicht (BMI)										
4	7.815	9.603	11.247	14.316	0.456	7.104	8.905	10.769	13.511	0.395
4(a)	6.349	7.770	8.872	11.548	0.387	5.727	7.442	8.416	11.013	0.327
4(b)	6.737	8.606	10.883	13.485	0.375	8.739	10.955	12.945	15.500	0.442

*Anmerkungen.*<sup>1</sup> Schwellenparameter = Punkt, an dem die Wahrscheinlichkeit für die Wahl der nächsthöheren Antwortkategorie größer wird als für die Wahl der vorherigen Antwortkategorie.<sup>2</sup> Steigung = Steigungskoeffizient der Regression.<sup>3</sup> Anzeige Balkendiagramm mit der geschlechtsspezifischen Merkmalsverteilung in der Population.<sup>4</sup> Die Ergebnisse für Körpergewicht in Kilogramm wurden aus Eichhorn (2016) entnommen. $t_1$  = erste Schwelle;  $t_2$  = zweite Schwelle;  $t_3$  = dritte Schwelle;  $t_4$  = vierte Schwelle.

Tabelle 2.3.11.

*Schwellenparameter ( $t^1$ ) und Steigung ( $s^2$ ) für Körpergewicht der Bedingung GV/GSV<sup>3</sup> für Frauen und Männer*

Merkmal und Item-Nr.	Frauen					Männer				
	$t_1$	$t_2$	$t_3$	$t_4$	$s$	$t_1$	$t_2$	$t_3$	$t_4$	$s$
Körpergewicht (kg) <sup>4</sup>										
4	6.657	8.398	9.941	12.857	0.139	6.249	8.489	9.956	13.226	0.111
4(a)	5.183	6.627	7.544	9.923	0.117	4.474	6.139	7.041	9.401	0.085
4(b)	6.747	8.710	10.847	13.456	0.128	4.656	6.706	8.156	10.455	0.083
Körpergewicht (BMI)										
4	7.546	9.305	10.866	13.878	0.431	8.508	10.824	12.415	15.912	0.458
4(a)	6.399	7.886	8.852	11.366	0.387	6.324	8.081	9.054	11.529	0.354
4(b)	7.419	9.385	11.531	14.154	0.390	6.884	9.058	10.663	13.229	0.364

*Anmerkungen.*<sup>1</sup> Schwellenparameter = Punkt, an dem die Wahrscheinlichkeit für die Wahl der nächsthöheren Antwortkategorie größer wird als für die Wahl der vorherigen Antwortkategorie.<sup>2</sup> Steigung = Steigungskoeffizient der Regression.<sup>3</sup> Gleichzeitige Anzeige Balkendiagramme mit der Gesamtmerkmalsverteilung und der geschlechtsspezifischen Merkmalsverteilung in der Population.<sup>4</sup> Die Ergebnisse für Körpergewicht in Kilogramm wurden aus Eichhorn (2016) entnommen. $t_1$  = erste Schwelle;  $t_2$  = zweite Schwelle;  $t_3$  = dritte Schwelle;  $t_4$  = vierte Schwelle.

Tabelle 2.3.12.

*Schwellenparameter ( $t^1$ ) und Steigung ( $s^2$ ) für Körpergewicht der Bedingung KV<sup>3</sup> für Frauen und Männer*

Merkmal und Item-Nr.	Frauen					Männer				
	$t_1$	$t_2$	$t_3$	$t_4$	$s$	$t_1$	$t_2$	$t_3$	$t_4$	$s$
Körpergewicht (kg) <sup>4</sup>										
4	6.067	8.050	9.860	12.778	0.138	7.370	9.023	10.368	13.881	0.118
4(a)	3.271	5.146	6.071	8.499	0.096	5.263	6.639	7.748	10.473	0.092
4(b)	4.887	6.898	9.032	11.611	0.102	7.181	8.635	10.351	14.029	0.108
Körpergewicht (BMI)										
4	7.394	9.422	11.304	14.384	0.450	8.705	10.382	11.755	15.515	0.440
4(a)	3.998	5.889	6.821	9.279	0.303	7.271	8.743	9.957	13.002	0.386
4(b)	6.054	8.123	10.414	13.281	0.342	8.255	9.739	11.503	15.266	0.397

*Anmerkungen.*<sup>1</sup> Schwellenparameter = Punkt, an dem die Wahrscheinlichkeit für die Wahl der nächsthöheren Antwortkategorie größer wird als für die Wahl der vorherigen Antwortkategorie.<sup>2</sup> Steigung = Steigungskoeffizient der Regression.<sup>3</sup> Anzeige keiner Merkmalsverteilung. Bei der Bedingung KV handelt es sich um die Kontrollbedingung.<sup>4</sup> Die Ergebnisse für Körpergewicht in Kilogramm wurden aus Eichhorn (2016) entnommen. $t_1$  = erste Schwelle;  $t_2$  = zweite Schwelle;  $t_3$  = dritte Schwelle;  $t_4$  = vierte Schwelle.

## 2.4 Diskussion Studie I

### 2.4.1 Zusammenfassung der Ergebnisse

Zielsetzung von Studie I war die Reanalyse vorhandener Daten, die im Rahmen der Masterarbeit der Autorin (Eichhorn, 2016) erhoben und analysiert wurden. Zentrale Fragestellungen waren, ob die Selbstbewertung abhängig vom Bildungsniveau unterschiedlich ausfällt und ob der BMI einen inkrementellen Beitrag für die Vorhersage des Antwortverhaltens bei Selbsteinschätzungen des Gewichts über das absolute Körpergewicht hinaus liefert.

Bei den Merkmalen Körpergröße, Körpergewicht und Einkommen zeigte sich bei beiden Geschlechtern ein Zusammenhang zwischen den realen Werten und der Antwortskala. Beim Merkmal Alter war dieser nur bei Frauen eingeschränkt erkennbar. Bei keinem Merkmal veränderte sich die Stärke des Zusammenhangs in Abhängigkeit vom Bildungsniveau.

Die Regressionsanalysen auf Basis des BMI fielen im Vergleich zu den Regressionsanalysen auf Basis des absoluten Körpergewichts bei Männern vergleichbar und bei Frauen in Bezug auf die Schwellenabstände ungünstiger aus. In Studie I lieferte der BMI innerhalb der Geschlechter keinen inkrementellen Beitrag für die Vorhersage des Antwortverhaltens bei Selbsteinschätzungen des Gewichts über das absolute Körpergewicht hinaus.

### 2.4.2 Interpretation und Implikationen der Ergebnisse

Beim Merkmal Alter fielen beim Bildungsniveau *Abitur* die Lagemaße der zentralen Tendenz über alle Bedingungen am geringsten aus. Somit bilden die Personen die jüngste Altersgruppe, die Abitur als höchsten Bildungsabschluss auswählten. Dies ist insofern nachvollziehbar, da bis zu 80 % der Abiturienten in Deutschland ein Studium anstreben (Schneider & Franke, 2014). Vermutlich handelt es sich bei der jüngsten Altersgruppe um Personen, die erst kürzlich eine höhere Schulbildung abgeschlossen haben und sich noch in weiterer Ausbildung befinden. Untermauert wird die Erklärung von der Beobachtung, dass diese Personengruppe im Mittel die niedrigsten Einkommenswerte aufweist. Dies könnte ein Indikator für eine gering vergütete Nebenbeschäftigung während eines Studiums sein.

Beim Merkmal Körpergewicht ergaben sich bei Frauen mit dem Bildungsniveau *Kein Abitur* die höchsten Lagemaße der zentralen Tendenz. So wiegen Frauen ohne Abitur im Mittel zwischen acht und zehn Kilo mehr als Frauen mit Hochschulabschluss. Bei Männern

ist ein solches Muster nicht erkennbar. Diese Beobachtung geht Hand in Hand mit der Studie zu Körpergewicht und assoziierten sozialen sowie ökonomischen Nachteilen bei Männern und Frauen von Sarlio-Lähteenkorva und Lahelma (1999). Sie konnten ein identisches, mehrfach repliziertes Muster zeigen (z. B. Gortmaker, Must, Perrin, Sobol & Dietz, 1993; Jeffery & French, 1996; Sargent & Blanchflower, 1994). Als Erklärung führen die Autoren an, dass Übergewicht mit Arbeitslosigkeit bzw. Langzeitarbeitslosigkeit assoziiert ist und ein höheres Risiko für Arbeitslosigkeit wiederum mit einem niedrigeren Bildungsniveau einhergeht. Weiter argumentieren die Autoren, dass der Zusammenhang von Übergewicht und niedrigem Bildungsniveau vor allem bei Frauen so offensichtlich ist, weil übergewichtige Frauen im Gegensatz zu übergewichtigen Männern am Arbeitsmarkt möglicherweise in höherem Maße diskriminiert werden (Sarlio-Lähteenkorva & Lahelma, 1999). Als Alternativerklärung könnte man heranziehen, dass niedriges Einkommen mit ungesunden Ernährungsgewohnheiten und Gewichtszunahme assoziiert ist (Jeffery & French, 1996) und dieser Zusammenhang eventuell durch ein geringeres Bildungsniveau moderiert wird. Dabei könnte Wissen über gesunde Ernährung ebenfalls als Teil höherer Bildung angesehen werden. Fakt ist, dass ein Zusammenhang zwischen Übergewicht und niedrigem Bildungsniveau bei Frauen bereits mehrfach beobachtet wurde und auch in Studie I klar ersichtlich ist. Über den Zusammenhang zwischen Ursache und Wirkung kann jedoch nur gemutmaßt werden, da eine experimentelle Untersuchung aus ethisch-moralischen Gründen nicht durchführbar ist. Selbst die eigens zu diesem Thema durchgeführten oben genannten Studien bleiben die Antwort auf die Frage der Kausalität schuldig.

Autoren testpsychologischer Fachbücher stellen immer wieder die Wichtigkeit der Berücksichtigung des Bildungsniveaus der Zielgruppe bei der Fragebogenkonstruktion heraus. Eine Nichtbeachtung könnte sich beispielsweise auf die Itemschwierigkeit oder Faktorstruktur des zu erfassenden Konstrukts auswirken (z. B. Bühner, 2011; Jonkisz, Moosbrugger & Brandt, 2012). Ein Grund für diese veränderten psychometrischen Eigenschaften könnte sein, dass weniger verständliche Items schwerer zu beantworten sind und dadurch gleichzeitig verbale Intelligenz miterfassen (Marsh, 1996). Dadurch wäre ihre Beantwortung nicht mehr unabhängig vom Bildungsniveau der Testpersonen, da ein höheres Bildungsniveau mit höherer verbaler Intelligenz einhergeht (Rowe, Jacobson & van den Oord, 1999). Demnach sollte die sprachliche Gestaltung von Testitems immer auf die

Zielgruppe zugeschnitten sein, über die eine Aussage getroffen werden soll (Bühner, 2011). In Studie I handelt es sich um Items, die Aussagen über einen möglichst breiten Personenkreis treffen sollen. Aus diesem Grund wurden die Items bewusst einfach und deutlich formuliert, sodass sie für eine große Personengruppe begreiflich und zugänglich sind. Obwohl die Verständlichkeit der Items strenggenommen erst mithilfe kognitiver Interviews überprüft werden müsste, wird postuliert, dass diese für die erwachsene Allgemeinbevölkerung keine größeren Verständnisprobleme bereiten sollten. Somit ist naheliegend, dass sich bei keinem Merkmal die Stärke des Zusammenhangs zwischen den realen Werten und der Antwortskala in Abhängigkeit des Bildungsniveaus verändert, da verbale Intelligenz bei vorliegender Einfachheit und Verständlichkeit der Items vermutlich eine untergeordnete Rolle spielt.

Zudem geben die Ergebnisse der Studie keinen Hinweis darauf, dass die Fähigkeit zur Selbsteinschätzung mit dem Bildungsniveau der Befragten in Zusammenhang steht. Diese Beobachtung geht einher mit den Befunden zur *Rosenberg-Skala* (Collani & Herzberg, 2003) bzw. zum *Job-Diagnostic-Survey* (Cordery & Sevastos, 1993). Bei beiden Testverfahren hatte das Bildungsniveau der Teilnehmer ebenfalls keinen Einfluss auf die Güte der Einschätzungen der Probanden.

Die in Eichhorn (2016) aufgeworfene Vermutung, dass Personen mit höherem Bildungsniveau die von den Balkendiagrammen transportierten Informationen eher in ihre Entscheidungen integrieren, weil sie Graphiken besser lesen und verstehen können (z. B. Glazer, 2011; Shah & Hoeffner, 2002), hat sich nicht bestätigt. Vermutlich sind Graphenverständnis und die Integration der generierten Informationen in einen darauffolgenden Entscheidungsprozess nicht zwangsläufig logisch miteinander verknüpft. So zeigt das Modell der begrenzten Rationalität, dass sich Menschen bei zunehmender Komplexität von Entscheidungssituationen aufgrund begrenzter kognitiver Fähigkeit generell nicht rationaler verhalten, sondern auf Heuristiken zurückgreifen (Simon, 1959). Dabei können auch Stimmungen und Gefühle als sogenannte Affektheuristik zur Urteilsbildung herangezogen werden (Finucane, Alhakami, Slovic & Johnson, 2000). Möglicherweise stößt die bei der Integration komplexer Informationen notwendige Kette kognitiver Prozesse an ein allgemeines Kapazitätslimit. In Folge dessen könnte, in Anlehnung an die Affektheuristik, auf ein automatisch und schnell entstehendes Gefühl, wie

beispielsweise das eigene Körpergefühl, zurückgegriffen werden. Dies muss nicht erst bewusst empfunden werden, sondern ermöglicht eine schnelle Entscheidung, was sich wiederum in begrenzt rationalem Verhalten in Form einer verzerrten Selbsteinschätzung widerspiegeln könnte.

Die Regressionsanalysen auf Basis des BMI fielen im Vergleich zu den Regressionsanalysen auf Basis des Körpergewichts bei Männern vergleichbar und bei Frauen in Bezug auf die Schwellenabstände ungünstiger aus. Bei allozentrischer Formulierung ergeben sich innerhalb der weiblichen Substichprobe keine äquidistanten Schwellen mehr. Somit bleiben die Relationen der realen Werte bei den Analysen auf Basis des BMI schlechter erhalten als bei Analysen auf Basis des absoluten Körpergewichts. Folglich scheint der BMI innerhalb der Geschlechter keine eigenständige Prognosekraft für die Vorhersage des Antwortverhaltens bei Selbsteinschätzungen des Gewichts zu haben, die über das absolute Körpergewicht hinausgeht.

Die in Eichhorn (2016) aufgeworfene Anregung, Analysen auf Basis des BMI durchzuführen, ist nachvollziehbar. So ist der BMI ein um die Körpergröße korrigiertes Maß des Körpergewichts (Birbaumer & Schmidt, 2010), das mit einer höheren Prädiktionsgüte einhergehen könnte. Allerdings ist der BMI als Maßzahl für die Bewertung des Körpergewichts eines Menschen nicht unumstritten, auch wenn die Berechnung einer plausiblen Logik folgt. Ein relativ hohes Körpergewicht und ein damit verbundener hoher BMI kann durch zahlreiche andere Faktoren, wie beispielsweise ausgeprägter Muskelmasse, größeren Durchmessern von Knochen und/oder Gelenken, höherer Knochendichte und markanterer Schulterbreite verursacht sein (Ahima & Lazar, 2013; Heymsfield, Scherzer, Pietrobelli, Lewis & Grunfeld, 2009; Krakauer & Krakauer, 2012; Nevill, Stewart, Olds & Holder, 2006). Diese Faktoren haben vor allem bei Sportlern eine hohe Relevanz. So konnte gezeigt werden, dass austrainierte Kraftsportler nur aufgrund ihrer Muskelmasse einen deutlich erhöhten BMI aufweisen, der jedoch keine Aussage über ein vorherrschendes Unter- bzw. Übergewicht treffen kann (Ruff, 2000). Darüber hinaus bleiben geschlechtsspezifische Unterschiede im Körperbau vom BMI unberücksichtigt, weswegen bei Frauen eine systematische Überschätzung und bei Männern eine systematische Unterschätzung des Körpergewichts durch den BMI stattfindet (Ahima & Lazar, 2013; Ruff, 2000).

### 2.4.3 Limitationen und Ausblick

Ein möglicher Kritikpunkt ist, dass die Stichprobe ein überwiegend hohes Bildungsniveau aufweist. So ist der Anteil an Personen, die mindestens Abitur haben, überproportional hoch (58.4 %) und der Anteil der Personen, die einen Hauptschulabschluss oder ein geringeres Bildungsniveau aufweisen, überproportional niedrig (12 %). Die Stichprobe kann nicht als repräsentativ für die Gesamtbevölkerung angesehen werden. Aussagen zum Einfluss des Bildungsniveaus auf das Item- und Graphenverständnis sowie auf die Fähigkeit zur Selbsteinschätzung können deshalb nicht gesichert getroffen werden. Für eine fundiertere Beantwortung dieser Fragestellung wäre eine Replikation der Studie an einer repräsentativen Stichprobe notwendig.

Die angeführte Kritik am BMI hat ihre Berechtigung und mögliche Erklärungen für die vorliegenden Ergebnisse können dort vermutet werden. Allerdings ist es wahrscheinlicher, dass eine Korrektur des Körpergewichts um die Körpergröße *innerhalb* der Geschlechter wenig zum Tragen kommt. Dies ist plausibel unter der Annahme, dass die mithilfe des BMI angedachte Korrektur bereits durch die getrennte Datenauswertung erreicht wird. Der Fall, dass sich Personen mit demselben Gewicht in ihrer Körpergröße stark unterscheiden und deshalb unterschiedlich bewerten, könnte vor allem zwischen den Geschlechtern auftreten. So ist die Verteilung der Körpergröße nach Geschlecht in der Population bimodal (GESIS - Leibniz-Institut für Sozialwissenschaften, 2015), wobei die Größenverteilung innerhalb der Geschlechter weniger divergiert als zwischen den Geschlechtern. Folglich könnte der BMI seine Stärke als Korrekturmaß vor allem dann zeigen, wenn eine gemeinsame, in Studie I nicht durchgeführte Datenauswertung für Männer und Frauen angestrebt wird.

### 2.4.4 Fazit

Bei der Reanalyse der Daten von Eichhorn (2016) hatte das Bildungsniveau bei keinem Merkmal Einfluss auf die Stärke des Zusammenhangs zwischen den realen Werten und der Antwortskala. Dies könnte darauf hinweisen, dass das Bildungsniveau bei derlei Fragestellungen eine untergeordnete Rolle spielt. Aufgrund der unter Abschnitt 2.4.3 beschriebenen Limitation sollte es bei weiterführenden Studien als möglicher Einflussfaktor dennoch kontrolliert werden.

Zudem konnte gezeigt werden, dass innerhalb der Geschlechter der BMI keine eigenständige Prognosekraft für die Vorhersage des Antwortverhaltens bei

Selbsteinschätzungen des Gewichts hat, die über das absolute Körpergewicht hinausgeht. Die durch den BMI angedachte Korrektur wurde möglicherweise durch die getrennte Datenauswertung bereits vorgenommen. Eine Korrektur des Körpergewichts um die Körpergröße mithilfe des BMI könnte deshalb vor allem bei einer für beide Geschlechter gemeinsamen Datenauswertung sinnvoll sein.



### 3. Studie II: Einfluss von Itemformulierung und Skalenformat auf das Skalenniveau

#### 3.1 Theorie Studie II

##### 3.1.1 Einführung

Bei der Formulierung von Fragebogenitems können selbst inhaltlich triviale Abwandlungen zu gravierenden Veränderungen in den Antworten der Befragten führen (Schwarz, 1999). Beispielsweise zeigten Schuman und Presser (1981) in ihren Untersuchungen, dass 48 % der Befragten antidemokratische Reden *nicht erlauben* würden, wohingegen nur 21 % diese *verbieten* würden. Selbst auf das zu erfassende Konstrukt hat die Itemformulierung maßgeblichen Einfluss. Clark und Watson (1995) zeigten, dass Items, die negative Stimmungen transportieren, wie beispielsweise „Ich Sorge mich über...“ oder „Ich bin verärgert über...“, Neurotizismus miterfassen. Werden solche affektgeladenen Items in einem Fragebogen verwendet, ist laut den Autoren die resultierende Skala vor allem ein Marker für Neurotizismus, unabhängig davon, welches Konstrukt eigentlich gemessen werden soll.

Überdies kann die numerische sowie verbale Skalenbeschriftung die Befragten in ihrem Antwortverhalten beeinflussen (z. B. Schwarz, Knäuper, Hippler, Noelle-Neumann & Clark, 1991). Selbst das Ausmaß, in dem die Befragten die Skala als äquidistant erleben, wird von der Art der numerischen Abstufung beeinflusst (Rohrmann, 1978). Daneben kann sich die Skalenbreite auf Reliabilität (z. B. Weng, 2004) und Validität (z. B. Hancock & Klockars, 1991) eines Fragebogens auswirken.

Die Qualität der erhobenen Daten wird also maßgeblich von der Qualität der Fragen bestimmt. Dabei kommt es nicht nur darauf an, *was* man fragt, sondern auch *wie* man fragt (Angleitner, John & Löhr, 1986). Für die Datenqualität nicht weniger kritisch ist die Wahl der passenden Antwortskala. Itemformulierung und Antwortskala stellen deshalb bei jeder Fragebogenkonstruktion Herausforderungen dar (Mummendey & Grau, 2014; Porst, 2014).

In der der Studie I zugrundeliegenden Arbeit (Eichhorn, 2016) bezogen sich die Itemaussagen durchgehend auf einen Pol des erfragten Merkmals (z. B. „Ich bin alt.“; „Ich bin groß.“; „Ich wiege viel.“). Die Antwortskala war sowohl verbal als auch numerisch verankert. Möglicherweise wurden die Befragten durch die auf einen Pol bezugnehmenden Formulierungen und die verbalen bzw. numerischen Verankerungen der Antwortskala in ihrem Antwortverhalten in eine bestimmte Richtung beeinflusst. Aus diesem Grund werden

in Studie II Itemformulierung und Antwortskala modifiziert. Die Items werden so verändert, dass ihre Formulierung keinen Bezug auf einen Pol des untersuchten Merkmals mehr aufweist. Die Beantwortung soll auf einer möglichst neutralen Antwortskala erfolgen. Dabei werden drei der gängigsten Skalenbreiten (fünfstufiges, vierstufiges und dichotomes Antwortformat) gegenübergestellt. Es wird erneut untersucht, ob sich die Selbstbewertungsperspektive als eine den Referenzrahmen der Selbstbewertung bestimmende Variable zeigt und ob das Bildungsniveau die Selbstbewertung beeinflusst. Im Mittelpunkt der Analysen steht die Frage, unter welcher Bedingung sich im Mittel die größte Annäherung an Erhalt von Ordnung, Unterschieden und Verhältnissen der realen Werte auf der Antwortskala ergibt (fünfstufiges und vierstufiges Antwortformat) bzw. inwieweit der Wendepunkt der logistischen Funktion mit den Lagemaßen der zentralen Tendenz korrespondiert (dichotomes Antwortformat). Der Datensatz von Studie II wurde inzwischen zur Veröffentlichung eingereicht (Hilbert et al., 2018).

### **3.1.2 Negativ gepolte und negativ formulierte Items**

Verschiedene Autoren empfehlen, Fragebögen so zu konzipieren, dass sie sowohl negativ als auch positiv formulierte Items enthalten bzw. sowohl negativ als auch positiv gepolte Items mit einbeziehen. Begründet wird dieses Vorgehen damit, dass so der Akquieszenz entgegengewirkt werden kann (z. B. Anastasi, 1982; Nunnally, 1978; Paulhus, 1991; Plieninger, 2017). Akquieszenz ist eine inhaltsunabhängige Zustimmungstendenz, die vor allem dann zum Tragen kommt, wenn die Zustimmung oder Ablehnung von Items nicht durch sozial erwünschtes Verhalten bestimmt wird (Diers, 1964). Neben der Zustimmungstendenz wird auch eine Ablehnungstendenz beschrieben, wobei eine Zustimmungstendenz häufiger beobachtet wird. Beide werden als relativ stabile Persönlichkeitseigenschaften angesehen (Couch & Keniston, 1960).

Zudem werden zu einem Themenbereich negativ formulierte Items konsistenter beantwortet als positiv formulierte Items. So seien die Befragten eher bereit, negativ formulierten Items zuzustimmen als positiv formulierte Items abzulehnen (Krebs & Matschinger, 1993).

Negativ gepolte Items sind so formuliert, dass die Zustimmungsrichtung entgegen der zugehörigen Skala verläuft. Niedrige Ausprägungen auf diesen Items entsprechen somit einer hohen Merkmalsausprägung (Bühner, 2011). Die negative Polung kann entweder

durch Verneinung (z. B. nicht glücklich), einer entsprechenden Vorsilbe (z. B. unglücklich) oder über das Antonym (z. B. traurig) erfolgen. In empirischen Untersuchungen werden negativ formulierte und negativ gepolte Items häufig nicht eindeutig unterschieden (Weijters & Baumgartner, 2012). Meistens wird die negative Polung durch Verneinung erreicht, weswegen vor allem Ergebnisse zu Items vorliegen, die gleichzeitig negativ gepolt und negativ formuliert sind (z. B. nicht glücklich).

Obwohl es nachvollziehbare Argumente gibt, Items unterschiedlicher Polung bzw. Formulierung zu verwenden, kann die Verwendung unterschiedlich gepolter Items nachteilige Effekte auf die psychometrischen Eigenschaften von Fragebögen haben. So können sie sowohl zu einer verschlechterten Konstruktvalidität (Meloni & Gana, 2001; Schriesheim, Eisenbach & Hill, 1991) als auch zu einer verringerten internen Konsistenz als Reliabilitätsmaß führen (Barnette, 2000; Hinz, Brähler, Geyer & Körner, 2003; Schriesheim et al., 1991; Schriesheim & Hill, 1981; Wang, Chen & Jin, 2015).

Durch unterschiedliche Itempolung kann die Faktorstruktur eines Fragebogens beeinträchtigt werden. So kann es sein, dass negativ gepolte Items einen eigenen Faktor bilden (Harvey, Billings & Nilan, 1985; Meloni & Gana, 2001; Miller & Cleary, 1993; Molina, Rodrigo, Losilla & Vives, 2014; Pilotte & Gable, 1990; Schmitt & Stults, 1985; Wang & Lin, 2011; für ein Review vgl. Weijters & Baumgartner, 2012). Konfirmatorische Faktorenanalysen zeigten, dass es sich hierbei um Methodenfaktoren handelt. Die negativ gepolten Items bilden einen artifiziellen Faktor, der auf die Unterschiede in der Itemformulierung zurückzuführen ist (Harvey et al., 1985; Höfling, Moosbrugger, Schermelleh-Engel & Heidenreich, 2011; Meloni & Gana, 2001). Die Umformulierung der negativ gepolten Items in positiv gepolte Items führte häufig zu einer eindimensionalen Struktur, bei der sich die Methodenfaktoren auflösten (Greenberger, Chen, Dmitrieva & Farrauggia, 2003; Idaszak & Drasgow, 1987; Wang & Lin, 2011). Ein Grund für diese veränderte Faktorstruktur kann nach Marsh (1996) sein, dass negativ gepolte Items schwerer zu verstehen sind und dadurch gleichzeitig verbale Intelligenz miterfasst wird. Seine Begründung stützt er darauf, dass mit steigenden verbalen Fähigkeiten der Befragten die Methodeneffekte negativer Polung abnehmen. Ferner zeigten Benson und Hocevar (1985), dass es schwerer ist, seine Zustimmung durch Widersprechen eines negativ gepolten Items auszudrücken als durch Zustimmung zu positiv gepolten Items.

Verschieden gepolte Items können zu unterschiedlichen Mittelwerten führen (McPherson & Mohr, 2005). Dies hat zur Folge, dass sich Summenwerte in eine bestimmte Richtung verzerren, falls die Anzahl positiver und negativer Items nicht ausbalanciert ist (Bühner, 2011).

Sobald sich bei der Bearbeitung von Fragebögen ein bestimmter Beantwortungsstil eingestellt hat, können sich laut Schmitt und Stults (1986) Personen schlechter an eine negative Itempolung anpassen. Dadurch werden negativ gepolte Items häufiger übersehen. Schon ein Anteil von 10 % der Versuchspersonen, die die negative Polung übersehen haben, kann für das Entstehen einer Faktorstruktur ausreichen, bei der ein Faktor die negativ gepolten Items repräsentiert (Schmitt & Stults, 1986). Die Bearbeitung von negativ gepolten Items erfordert laut Swain, Weathers und Niedrich (2008) einen höheren kognitiven Aufwand, weshalb dabei fehlerhaftes Antwortverhalten um ca. 20 % höher zu verzeichnen ist als bei nicht negativ gepolten Items. Vor allem vereinzelte negativ gepolte Items nach vielen positiv gepolten Items begünstigen Beantwortungsfehler (Drolet & Morrison, 2001).

Bassili und Scott (1996) zeigten, dass die Verwendung negativ gepolter bzw. negativ formulierter Items zu signifikant längeren Reaktionszeiten bei den Befragten führt. Sie werteten die längeren Reaktionszeiten als Kennzeichen unzulänglich formulierter Items, deren Beantwortung mit erhöhter kognitiver Belastung einhergeht.

Unterschiede in der Itempolung beeinflussen auch die Itemkorrelation. Items gleicher Polung korrelieren höher als Items unterschiedlicher Polung (Hinz et al., 2003; Miller & Cleary, 1993). Die Korrelation zwischen positiv und negativ gepolten Items fällt nur moderat aus (Höfling et al., 2011; Meloni & Gana, 2001). Geht man davon aus, dass unterschiedlich gepolte Items grundsätzlich dasselbe Merkmal messen, wäre eine wesentlich höhere Korrelation zu erwarten. Auch die Korrelation der Summenwerte negativ bzw. positiv gepolter Items mit Fremdbeurteilungen fällt unterschiedlich hoch aus (Bühner, 2011). Es ist also fraglich, ob negativ formulierte bzw. negativ gepolte Items dasselbe Konstrukt erfassen wie positiv formulierte bzw. positiv gepolte Items. Möglicherweise wird das zu erfassende Konstrukt durch die Umpolung verändert. Dies würde Mummendeys These bestätigen, die besagt, dass es fast unmöglich ist „zu einer bestimmten Formulierung eine vollständig bedeutungsgleiche, nur sozusagen mit anderem Vorzeichen versehene Formulierungs-Umkehrung zu erhalten“ (2003, S. 66).

### 3.1.3 Antwortskalen mit festen Intervallen

Bei der Konzeption eines Fragebogens ist die Wahl der Antwortskala ein wichtiger Arbeitsschritt, bei dem mehrere Aspekte zu berücksichtigen sind. Zunächst kann die Anzahl der Antwortstufen, die sogenannte Skalenbreite, die psychometrischen Eigenschaften des Fragebogens beeinflussen. Studien zeigen, dass sich mit zunehmender Skalenbreite die Reliabilität (z. B. Lozano, García-Cueto & Muñiz, 2008; Maydeu-Olivares, Kramp, García-Forero, Gallardo-Pujol & Coffman, 2009; Oaster, 1989; Preston & Colman, 2000; Weng, 2004) und die Validität (z. B. Hancock & Klockars, 1991; Lozano et al., 2008; Preston & Colman, 2000) des Fragebogens erhöhen kann. Allerdings scheint dieser Anstieg nicht linear zu verlaufen. So berichtete Birkett (1986) die höchste interne Konsistenz bei einer sechsstufigen Skala im Vergleich zu einer dichotomen und 14-stufigen Skala. Finn (1972) erzielte die höchste Reliabilität bei einer siebenstufigen Skala verglichen mit einer drei-, fünf- und neunstufigen Skala. Cicchetti, Showalter und Tyrer (1985) untersuchten den Zuwachs an Reliabilität bei ein- bis hundertstufigen Skalen. Dabei ergab sich die höchste Reliabilität bei einer siebenstufigen Skala. Preston und Colman (2000) untersuchten Reliabilität und Validität von Skalenbreiten zwischen zwei und elf Stufen. Die höchste Reliabilität und Validität ergab sich ebenfalls bei einer siebenstufigen Skala.

In Frage gestellt werden die Ergebnisse von Drake, Hargraves, Lloyd, Gallagher und Cleary (2014). Sie konnten in ihrer Studie keinen Reliabilitätsunterschied zwischen einer vierstufigen und einer sechsstufigen Skala ausmachen. Maydeu-Olivares und Kollegen (2009) konnten den Zusammenhang von Skalenbreite und psychometrischen Eigenschaften nur für die Reliabilität, jedoch nicht für die Validität zeigen. Matell und Jacoby (1971) konnten keinen Zusammenhang zwischen Skalenbreite und Reliabilität sowie Skalenbreite und Validität finden. Allerdings sind deren Ergebnisse aufgrund der geringen Stichprobengröße ( $n = 20$ ) und einer möglicherweise dadurch verursachten unzureichenden statistischen Power wenig belastbar.

Trotz uneindeutiger Studienlage scheinen siebenstufige Skalen hinsichtlich psychometrischer Merkmale optimal zu sein, weshalb sie oft empfohlen werden (vgl. Cicchetti et al., 1985; Finn, 1972; Hilbert, Küchenhoff, Sarubin, Nakagawa & Bühner, 2016; Schwarz et al., 1991; für ein Review vgl. Cox III, 1980). In der Praxis werden am häufigsten

fünfstufige Antwortskalen verwendet (Lozano et al., 2008). Dies entspricht der von Likert (1932) empfohlenen Skalenbreite, wobei er die Zahl fünf nie psychometrisch begründete.

Die optimale Skalenbreite hängt auch von der Abstraktionsfähigkeit der Befragten ab. So können zu breite Skalen schnell zu intellektueller Überforderung bei den Befragten führen, da sie die Skalenpunkte nicht mehr voneinander abgrenzen oder sinnvoll definieren können (Krosnick & Presser, 2010; Porst, 2014; Rohrmann, 1978). Außerdem suggerieren breite Skalen eine Scheinpräzision, wenn beispielsweise Skalenpunkt 14 von Skalenpunkt 15 nicht sinnvoll unterschieden werden kann. Zu enge Skalen können Widerwillen bei den Befragten auslösen, da sie keine befriedigende Möglichkeit zur ausreichenden Differenzierung bieten (Porst, 2014).

Die Antwortskala kann *gerade* sein, d. h. eine gerade Anzahl an Antwortstufen aufweisen oder *ungerade* sein, d. h. eine ungerade Anzahl an Antwortstufen haben. Beim Einsatz einer ungeraden Skala wird automatisch eine Mittelkategorie zur Verfügung gestellt. Dies kann fehlerhaftes Antwortverhalten und extreme Antwortstile reduzieren (Weijters, Cabooter & Schillewaert, 2010). Laut Marsh und Parducci (1978) werden Skalen mit Mittelkategorie von den Befragten eher als äquidistant wahrgenommen als Skalen ohne Mittelkategorie. Allerdings sind Antworten auf dem Skalenmittelpunkt schwer zu interpretieren (Mummendey & Grau, 2014). Außerdem wird die Mittelkategorie von den Befragten häufig als Fluchtkategorie verwendet, wenn der Befragte die Auskunft verweigern (Bühner, 2011; Mummendey & Grau, 2014; Porst, 2014) oder kognitive Anstrengungen bei der Fragebogenbearbeitung reduzieren möchte (Krosnick, 1991). Garland (1991) konnte zeigen, dass gerade Skalen sozial erwünschtes Antwortverhalten reduzieren können. Jedoch zwingen gerade Skalen die Befragten zu einer Entscheidung und nehmen ihnen die Möglichkeit, sich bewusst auf einer Mittelposition einzuordnen. Diese Freiheitseinkerbung kann wiederum zu einer Nichtbeantwortung oder willkürlichen Fragenbeantwortung führen (Mummendey & Grau, 2014; Porst, 2014). Jeder der beiden Skalentypen kann also zu Datenverzerrungen führen. Deshalb gibt es Empfehlungen sowohl für eine gerade Skala (z. B. Bühner, 2011) als auch für eine ungerade Skala (z. B. Mummendey & Grau, 2014; Sturgis, Roberts & Smith, 2014).

Bei *verbalisierten* Skalen ist jeder Skalenpunkt mit einer verbalen Umschreibung versehen. Bei *endpunktbenannten* Skalen sind nur die beiden Skalenpole verbal verankert.

Das Ziel ist dabei immer die Konzeption möglichst äquidistanter Antwortstufen, um die Skala dem Intervallskalenniveau anzunähern (Mummendey & Grau, 2014; Porst, 2014). Allerdings bilden verbalisierte Skalen lediglich eine Rangordnung der Antwortmöglichkeiten ab (Porst, 2014). Zudem wird das Finden von angemessenen und semantisch gleichabständigen Begriffen mit zunehmender Skalenbreite immer schwieriger (Rohrmann, 1978). Aus diesem Grund wird verbalisierten Skalen meist nur Ordinalskalenniveau zugeschrieben (Porst, 2014). Weijters und Kollegen (2010) zeigten, dass verbalisierte Skalen zwar Akquieszenz begünstigen, jedoch extreme Antwortstile und fehlerhaftes Antwortverhalten reduzieren. Auch können verbalisierte Skalen verglichen mit endpunktbenannten Skalen eine höhere Reliabilität aufweisen (Krosnick & Berent, 1990; Weng, 2004). Ferner sind sie aufgrund eindeutiger Vorgaben bei Befragten beliebter (Porst, 2014; Rohrmann, 1978). Endpunktbenannte Skalen entsprechen eher der Idee intervallskalierter Skalen. So ergibt sich durch Vorgabe gleichabständiger Antwortkästchen oder Ziffern eine äquidistant empfundene Skala (Mummendey & Grau, 2014; Porst, 2014; Rohrmann, 1978). Die Befragten müssen die Bewertungsskala ohne verbale Anker mental abbilden und den jeweiligen Kategorien Bedeutungen zuordnen. Dies kann zu beliebiger Interpretation der Skalenpunkte und/oder Ambiguität führen (Mummendey & Grau, 2014; Porst, 2014; Weijters et al., 2010). Außerdem geht die Verwendung endpunktbenannter Skalen in der Regel mit einer höheren kognitiven Belastung einher, was in vermehrt fehlerhaftem Antwortverhalten resultieren kann (Weijters et al., 2010).

Antwortskalen sind häufig numerisch verankert, entweder anstelle von oder zusammen mit verbalen Verankerungen. Dabei verläuft die Nummerierung in der Regel von links nach rechts. Dies hat beispielsweise den Vorteil, dass daraus bestimmte Mittelwerte intuitiv nachvollziehbar sind (Porst, 2014). Numerisch verankerte Skalen können den Eindruck äquidistanter Antwortkategorien unterstützen (Mummendey & Grau, 2014), wobei die Art der numerischen Verankerung dabei zu berücksichtigen ist. So werden Skalen mit positiver Bezifferung (z. B. 1/2/3/4/5) eher als äquidistant wahrgenommen als Skalen mit bipolarer Bezifferung (z. B. -2/-1/0/+1/+2) (Rohrmann, 1978). Schwarz und Kollegen (1991) zeigten, dass bei identischer verbaler Verankerung und Skalenbreite negativ bezifferte Antwortkategorien signifikant weniger verwendet werden als die korrespondierenden positiv bezifferten Antwortkategorien. In einer darauffolgenden Studie konnte dieses

Antwortmuster repliziert werden (Schwarz & Hippler, 1995). Mummendey und Grau (2014) erklären dieses häufig beobachtete Phänomen damit, dass sich Befragte nicht mit negativen Zahlen in Verbindung bringen wollen. Eine Alternativerklärung ist, dass die Art der numerischen Verankerung die semantische Bedeutung der Kategorien beeinflusst. So wird beispielsweise bei Erfolgseinschätzungen die Ziffer 0 als Abwesenheit von Erfolg angesehen, wohingegen die Ziffer -5 als Anwesenheit von Misserfolg aufgefasst wird (Schwarz & Hippler, 1995).

Aufgrund der Vielzahl an sich oft widersprechenden Forschungsergebnissen zu Antwortskalen sowie unzähliger Einflussfaktoren, die bei der Beantwortung von Fragebögen wirken, zeichnet sich keine klare Empfehlung für *die* optimale Antwortskala ab. Porst bringt es vermutlich auf den Punkt, wenn er sagt: „Egal für welche Skala Sie sich letztendlich entscheiden: Sie müssen davon ausgehen, dass Sie immer einen Fehler machen“ (2014, S. 94).

#### **3.1.4 Rationale**

Unter Abschnitt 3.1.2 und 3.1.3 wurde verdeutlicht, dass Itempolung, Skalenbreite sowie Skalenbeschriftung das Antwortverhalten der Befragten und damit deren Selbstbewertungsprozess beeinflussen können. In der der Studie I zugrundeliegenden Arbeit (Eichhorn, 2016) bezogen sich alle Itemaussagen auf einen Pol des erfragten Merkmals (z. B. „Ich bin alt.“ oder „Ich wiege viel.“). Die Antworten mussten auf einer fünfstufigen, verbal und numerisch verankerten Likert-Skala (Likert, 1932) gegeben werden. Möglicherweise wurden dadurch die Befragten in ihrem Antwortverhalten in eine bestimmte Richtung beeinflusst. Aufgrund dieser Überlegung werden in Studie II Itemformulierungen verwendet, die sich nicht auf einen bestimmten Pol beziehen (z. B. „Wie schätzen Sie Ihr Alter ein?“). Die Antwortskala wird an die veränderte Itemformulierung angepasst. Dabei werden drei verschiedenstufige bipolare Skalen (fünf- und vierstufige sowie dichotome Skala) gegenübergestellt, bei denen die Pole jeweils mit *niedrig* bzw. *hoch* beschriftet sind. Es wird keine weitere numerische oder verbale Verankerung vorgenommen. Außerdem wird exploriert, ob sich die Selbstbewertungsperspektive erneut als eine den Referenzrahmen der Selbstbewertung bestimmende Variable zeigt und ob die Selbstbewertung abhängig vom Bildungsniveau unterschiedlich ausfällt.



Die Fragestellungen von Studie II basieren auf denselben Annahmen wie in Eichhorn (2016) bzw. Studie I (vgl. Abschnitt 2.1.4). Untersucht wird, bis zu welchem Grad Ratingskalen metrische Informationen wiedergeben können, d. h. unter welcher Bedingung sich im Mittel die größte Annäherung an Erhalt von Ordnung, Unterschieden und Verhältnissen der realen Werte auf der Antwortskala zeigt (fünfstufiges und vierstufiges Antwortformat) bzw. inwieweit der Wendepunkt der logistischen Funktion mit den Lagemaßen der zentralen Tendenz übereinstimmt (dichotomes Antwortformat).

Die Realisierung erfolgt in einem zweifaktoriellen gemischten Design. Zunächst werden die Merkmale Alter, monatliches Nettoeinkommen, Körpergröße und Körpergewicht von allen Teilnehmern erhoben. Sie stellen die realen Werte der untersuchten Merkmale dar.

*Faktor mit Messwiederholung.* Alle Teilnehmer sollen sich aus zwei verschiedenen Perspektiven in den zuvor erhobenen Merkmalen selbst bewerten. Dabei werden, wie bereits in Studie I, die Konzepte des allozentrischen und egozentrischen Referenzrahmens aufgegriffen und auf die Itemformulierung angewendet. Durch zwei unterschiedliche Formulierungen (z. B. „Wie schätzen Sie Ihr Alter ein?“ und „Wie schätzen andere Ihr Alter ein?“) soll ein innerer Perspektivenwechsel bei der Testperson erreicht werden (vgl. Klatzky, 1998; Mandel, 2003).

*Faktor ohne Messwiederholung.* Die Selbstbewertung wird entweder auf einer fünfstufigen (5PS), vierstufigen (4PS) oder dichotomen Antwortskala (2PS) vorgenommen. In drei Versionen des Fragebogens beantworten die Probanden dieselben Items mit Antwortskalen unterschiedlicher Breite.

Auf Basis vorhergehender Überlegungen ergeben sich folgende Forschungsfragen:

1. Bleiben die Relationen der realen Werte unterschiedlich gut erhalten in Abhängigkeit von Itemformulierung und Antwortskala (Vergleich der Bedingung 5PS mit den Ergebnissen aus Eichhorn, 2016)?
2. Bei welcher Skalenbreite bleiben die Relationen der realen Werte am besten erhalten?
3. Kann durch unterschiedliche Itemformulierung ein Perspektivenwechsel in der Selbstbewertung erreicht werden, der dazu führt, dass die Relationen der realen Werte zwischen den Bedingungen im Mittel unterschiedlich gut erhalten bleiben (5PS und 4PS) bzw. dass die Lagemaße der zentralen Tendenz mit dem Wendepunkt der logistischen Funktion unterschiedlich gut korrespondieren (2PS)?

4. Fällt die Selbstbewertung in Abhängigkeit vom Bildungsniveau unterschiedlich aus?

### **3.2 Methoden Studie II**

#### **3.2.1 Präregistrierung**

Studie II wurde im Open Science Framework vor Beginn der Datenerhebung präregistriert. Unter folgendem Link ist die Präregistrierung einsehbar:

[https://osf.io/gs9mw/?view\\_only=4e1823649f7744489ec9ad787f068cd6](https://osf.io/gs9mw/?view_only=4e1823649f7744489ec9ad787f068cd6)

#### **3.2.2 Stichprobe**

Für die Untersuchung konnten insgesamt 1 903 Teilnehmer geworben werden. Davon wurden 134 Probanden ausgeschlossen, weil sie den Fragebogen nicht vollständig bearbeitet hatten. Es ergab sich somit eine Gesamtstichprobe von 1 769 Teilnehmern im Alter von 18 bis 77 Jahren ( $M = 32$ ,  $SD = 10.8$ ), darunter 1 405 Frauen (79.4 %) und 364 Männer (20.6 %). Beim Bildungsniveau gaben 33.9 % der Befragten an, dass sie keine allgemeine Hochschulreife haben und 66.1 % der Personen, dass sie allgemeine Hochschulreife haben.

Die Gesamtstichprobe bestand aus drei unabhängigen Substichproben, die sich auf die Bedingungen 5PS, 4PS und 2PS verteilten. Die drei Bedingungen werden unter Abschnitt 3.2.3 genauer erläutert. Tabelle 3.2.1. zeigt die Verteilung von Geschlecht, Bildungsniveau und Alter innerhalb der drei Bedingungen. Dabei ist ersichtlich, dass es sich um weitestgehend homogene Teilstichproben handelt. Aufgrund der linkssteilen Altersverteilung und Ausreißerwerten wird der Median anstatt des Mittelwerts als Maß der zentralen Tendenz berichtet.

Die Anwerbung der Probanden erfolgte über das soziale Netzwerk Facebook, über private E-Mail-Verteiler der Autorin sowie über den Studien-Informations-Service der Ludwig-Maximilians-Universität München (LMU). Die Teilnahme war freiwillig und ohne Vergütung.

Tabelle 3.2.1.

Verteilung von Geschlecht, Bildungsniveau und Alter innerhalb der Bedingungen 5PS<sup>1</sup>, 4PS<sup>2</sup> und 2PS<sup>3</sup>

Merkmal		5PS <sup>1</sup>		4PS <sup>2</sup>		2PS <sup>3</sup>	
		<i>n</i>	%	<i>n</i>	%	<i>n</i>	%
Geschlecht							
	Frauen	458	77.8	481	81.0	466	79.5
	Männer	131	22.2	113	19.0	120	20.5
Bildungsniveau							
	Kein Abitur	201	34.1	190	32.0	209	35.7
	Abitur	388	65.9	404	68.0	377	64.3
		<i>Md</i>	<i>SD</i>	<i>Md</i>	<i>SD</i>	<i>Md</i>	<i>SD</i>
Alter		28	11.2	29	10.2	29	11.0
Gesamtteilnehmerzahl		589		594		586	

Anmerkungen.

<sup>1</sup> Fünfstufige Antwortskala.<sup>2</sup> Vierstufige Antwortskala.<sup>3</sup> Dichotome Antwortskala.*n* = Anzahl; % = prozentualer Anteil innerhalb der Teilstichprobe; *Md* = Median; *SD* = Standardabweichung.

### 3.2.3 Material

Zur Untersuchung der Fragestellung wurde ein Fragebogen erstellt, in dem zunächst Alter, Körpergröße, Körpergewicht und monatliches Nettoeinkommen der Teilnehmer numerisch und ganzzahlig abgefragt wurden. Sie bildeten die realen Werte der untersuchten Merkmale ab. Analog zu Studie I sind davon die wahren Werte eines Merkmals im Sinne der KTT abzugrenzen (vgl. Abschnitt 2.2.2). In Studie II wird sich ebenfalls nicht auf die Definition der KTT bezogen, weshalb die sprachliche Abgrenzung beibehalten wurde. Zudem wurden mithilfe vorgegebener Auswahlmöglichkeiten Bildungsniveau (*Keine Allgemeine Hochschulreife* oder *Allgemeine Hochschulreife (Allgemeines Abitur)*) und Geschlecht (*weiblich* oder *männlich*) erhoben. In Studie II wurde das Bildungsniveau binär erhoben, da sich in Studie I zeigte, dass einzelne Kategorien kaum besetzt waren (z. B. *Kein Schulabschluss* oder *Hauptschulabschluss*) und deshalb die angesetzten fünf Kategorien in drei Kategorien zusammengefasst werden mussten (vgl. Abschnitt 2.2.4).

Auf den folgenden vier Fragebogenseiten wurden die Probanden aufgefordert eine Selbsteinschätzung bezüglich der vier Merkmale Alter, Körpergröße, Körpergewicht sowie

monatliches Nettoeinkommen vorzunehmen. Dabei wurde für alle Merkmale jeweils eine Formulierung mit egozentrischer Perspektive (z. B. „Wie schätzen Sie Ihr Alter ein?“) und mit allozentrischer Perspektive (z. B. „Wie schätzen andere Ihr Alter ein?“) vorgegeben (vgl. Klatzky, 1998; Mandel, 2003). Eine Übersicht aller Items findet sich in Anhang B (Tabelle B.1.).

Unter drei Bedingungen ohne Messwiederholung sollten die Teilnehmer die Items entweder auf einer fünfstufigen (5PS), vierstufigen (4PS) oder dichotomen Skala (2PS) beantworten. Dabei wurden bei jeder Skala nur die Skalenpole mit den Worten *niedrig* bzw. *hoch* beschriftet. Darüber hinaus wurde keine weitere verbale oder numerische Verankerung vorgenommen. Die Teilnehmer sollten von Itemformulierungen, die einen bestimmten Pol ansprechen (z. B. „Ich bin alt“) und/oder vorgegebenen Skalenbeschriftungen (z. B. Ziffern) möglichst unbeeinflusst bleiben.

Zum besseren Verständnis der Skalen wurden zu Beginn jeder Fragebogenseite folgende Bearbeitungshinweise gegeben: *Wählen Sie für Ihre Einschätzung einen Wert auf der Skala von „niedrig“ bis „hoch“. Zwischenwerte können entsprechend gewählt werden.* (fünfstufige und vierstufige Skalen) bzw. *Wählen Sie für Ihre Einschätzung entweder den Wert "niedrig" oder "hoch".* (dichotome Skala). Die gewählten Antwortkategorien gingen als die beobachteten Werte einer Person beim jeweils abgefragten Merkmal in die Datenanalyse ein.

### **3.2.4 Erhebung**

Die Datenerhebung erfolgte vom 08. Mai bis 18. Juni 2017 in Form eines Onlinefragebogens. Dieser wurde mit der Befragungssoftware SoSci Survey (Leiner, 2014) erstellt und den Teilnehmern auf [www.soscisurvey.de](http://www.soscisurvey.de) zur Verfügung gestellt. Der Fragebogen konnte von heimischen oder mobilen internetfähigen Geräten aus bearbeitet werden. Eine Kontrolle äußerer Bedingungen war nicht möglich.

Voraussetzung für die Teilnahme war ein Mindestalter von 18 Jahren, da es sich hierbei um eine Folgestudie von Studie I handelt und die Stichproben vergleichbar sein sollten (aufgrund der angezeigten Merkmalsverteilungen wurde in Studie I das Mindestalter auf 18 Jahre festgelegt). Überdies gab es keine Ein- bzw. Ausschlusskriterien.

Zunächst wurden alle Teilnehmer mithilfe einer vorgeschalteten Textseite über die Freiwilligkeit und Anonymität der Teilnahme aufgeklärt sowie über den Zweck der

Untersuchung und die zu erhebenden Daten informiert. Durch Klicken des „Weiter“-Buttons konnten die Befragten ihr Einverständnis mit den Durchführungsbedingungen bestätigen.

Auf der darauffolgenden Seite wurden die Merkmale Alter, Bildungsniveau, Körpergröße, Körpergewicht, monatliches Nettoeinkommen und Geschlecht aller Probanden erhoben. Sobald alle Fragen beantwortet und der „Weiter“-Button geklickt wurde, aktivierte sich eine Urnenfunktion ohne Zurücklegen. Diese sorgte für eine zufallsgesteuerte Zuweisung der Teilnehmer zu einer der drei Bedingungen ohne Messwiederholung. In der Urne befanden sich die Fragebogenversionen der drei Bedingungen 5PS, 4PS und 2PS. War die Urne geleert, füllte sie sich automatisch wieder mit den drei Fragebogenversionen und startete von vorne. Die Urnenfunktion stellte sicher, dass sich annähernd vergleichbare Experimentalgruppen ergaben (Bosch, 2011).

Auf den weiteren vier Seiten des Fragebogens folgten jeweils zwei Items zur Selbsteinschätzung aus zwei verschiedenen Perspektiven (je Seite ein Merkmal). Abschließend wurde die erfolgreiche Speicherung der Daten bestätigt und sich für die Teilnahme bedankt. Ein kompletter Onlinefragebogen in der Version 4PS findet sich in Form von Screenshots in Anhang B (Abbildungen B.1. bis B.7.).

### **3.2.5 Analysen**

Die Datenanalyse erfolgte mit der Open Source Statistiksoftware R (R Core Team, 2017). Für die Berechnung wurden die R-Pakete „foreign“ (R Core Team, 2016), „dplyr“ (Wickham & Francois, 2016), „ordinal“ (Christensen, 2015), „MASS“ (Venables & Ripley, 2002) und „brant“ (Schlegel & Steenbergen, 2017) verwendet. Die Diagramme wurden mit dem R-Paket „ggplot2“ (Wickham, 2009) oder Microsoft Excel erstellt.

Zur besseren Vergleichbarkeit mit den Ergebnissen in Eichhorn (2016) wurden alle Analysen ebenfalls getrennt für Frauen und Männer durchgeführt. Beim Merkmal Einkommen wurden die Regressionsanalysen mit skalierten Werten, nämlich Tausend-Euro-Werten (kEUR), durchgeführt. Dabei wurde das angegebene monatliche Nettoeinkommen jedes Teilnehmers durch tausend dividiert.

Beim verwendeten Fragebogen gaben die Teilnehmer entweder auf einer fünfstufigen, vierstufigen oder dichotomen Antwortskala ihre Einschätzung zu den jeweiligen Itemaussagen an. Die Antwortalternativen wurden wie folgt kodiert: Bei der fünfstufigen

Antwortskala *niedrig* mit der Ziffer 1, *hoch* mit der Ziffer 5 und die dazwischenliegenden Kategorien aufsteigend mit den Ziffern 2, 3 und 4; bei der vierstufigen Antwortskala *niedrig* mit der Ziffer 1, *hoch* mit der Ziffer 4 und die dazwischenliegenden Kategorien aufsteigend mit den Ziffern 2 und 3; bei der dichotomen Antwortskala mussten aufgrund der Modellanforderungen für logistische Regressionen die Antwortalternativen als Dummy-Variable codiert werden, nämlich *niedrig* mit der Ziffer 0 und *hoch* mit der Ziffer 1 (Agresti, 2013).

Anschließend wurden die Daten deskriptiv-statistisch ausgewertet. Zunächst wurden Mittelwert, Standardabweichung, Median, kleinster Wert und größter Wert der realen Werte für jedes der Merkmale Alter, Körpergröße, Körpergewicht und Einkommen jeweils für Frauen und Männer ermittelt und tabellarisch gegenübergestellt. Zu den Häufigkeiten der realen Werte wurde für jedes Merkmal ein Histogramm erstellt. Darin wurden jeweils beide Geschlechter getrennt dargestellt. Des Weiteren wurden Mittelwert, Median und Standardabweichung für die vier erhobenen Merkmale innerhalb jeder Bedingung (5PS, 4PS und 2PS) getrennt nach Bildungsniveau ermittelt und tabellarisch gegenübergestellt.

Für die beobachteten Werte wurden Mittelwert, Standardabweichung und Median für alle sechs Bedingungen jeweils für Frauen und Männer ermittelt und tabellarisch gegenübergestellt.

Zur Untersuchung von Zusammenhängen zwischen gewählter Antwortkategorie und realem Wert einer Person wurden für alle Merkmale und Bedingungen Spearman-Korrelationen (5PS und 4PS) bzw. punktbiseriale Korrelationen (2PS) berechnet. Es handelt sich dabei um ergänzende explorative Analysen, die über die präregistrierten Analysen hinaus gehen.

Zur Untersuchung der Beziehung zwischen den realen Werten und der Antwortskala wurden für die Bedingungen 5PS sowie 4PS Proportional Odds Regressionen und für die Bedingung 2PS logistische Regressionen durchgeführt. Bei Proportional Odds Regressionen wird die Chance für die Wahl einer Antwortkategorie ins Verhältnis gesetzt zur Chance für die Wahl der vorhergehenden Antwortkategorie. Der Punkt, an dem die Wahl der nächst höheren Kategorie wahrscheinlicher wird als die Wahl der vorherigen Kategorie, wird als Schwelle oder Schwellenparameter bezeichnet. Bei der logistischen Regression kennzeichnet der Wendepunkt der logistischen Funktion die Stelle, an der die Wahl der

Kategorie *hoch* wahrscheinlicher wird als die Wahl der Kategorie *niedrig* (Agresti, 2010; Tutz, 2012).

Für die Bedingungen 5PS und 4PS wurden die Lage der Schwellenparameter (rot) und für die Bedingung 2PS die logistischen Funktionen (rot) graphisch dargestellt. In jedem Diagramm wurden Mittelwert (blau) und Median (grün) der realen Werte aus der jeweiligen Teilstichprobe des nicht-messwiederholten Faktors abgebildet. Anhand der Diagramme wurden Unterschiede in den Abständen der Schwellenparameter zwischen den Bedingungen untersucht (5PS und 4PS). Außerdem wurde überprüft, ob der Mittelwert der realen Werte in der Mittelkategorie der Skala liegt (5PS) bzw. ob der Mittelwert mit der zweiten Schwelle (4PS) respektive mit dem Wendepunkt der logistischen Funktion (2PS) korrespondiert. Je besser die Informationen über die Intervalle der realen Werte erhalten bleiben, desto stärker sollte eine Annäherung an äquidistante Schwellen stattfinden (5PS und 4PS) und der Mittelwert der realen Werte in der Mittelkategorie liegen (5PS) bzw. mit der zweiten Schwelle (4PS) respektive mit dem Wendepunkt (2PS) korrespondieren.

Ergänzend wurden für alle Bedingungen die Schwellenparameter mit den entsprechenden Steigungskoeffizienten tabellarisch berichtet. Der Steigungskoeffizient ist ein Maß für Stärke und Richtung des Zusammenhangs zwischen den realen Werten und der Ratingskala. Ein positiver Wert zeigt an, dass bei zunehmend höherer Merkmalsausprägung eine zunehmend höhere Antwortkategorie gewählt wird. Es wird umso schneller eine höhere Antwortkategorie gewählt, je höher der Wert des Steigungskoeffizienten ausfällt (Agresti, 2010; Tutz, 2012).

Zur Überprüfung der Annahme äquidistanter Schwellen wurde zu jeder Proportional Odds Regression ein Likelihood Ratio Test durchgeführt. Dabei wird ein weniger restriktives Modell mit flexiblen Schwellen gegen ein restriktiveres Modell mit äquidistanten Schwellen getestet. Ein signifikanter Likelihood Ratio Test zeigt an, dass das Modell mit flexiblen Schwellen die Daten signifikant besser abbildet und die Annahme äquidistanter Schwellen verworfen werden muss (Norušis, 2012).

Zur Überprüfung der Frage, ob die Abbildung der Relationen der realen Werte auf der Antwortskala in Abhängigkeit von Itemformulierung (Itemformulierung, die nicht auf einen Pol des erfragten Merkmals Bezug nimmt vs. Itemformulierung, die sich auf einen Pol des erfragten Merkmals bezieht) und Antwortskala (endpunktbenannte Antwortskala ohne

numerische und verbale Verankerung vs. numerisch und verbal verankerte Likert-Skala) divergiert, wurde Bedingung 5PS mit den Ergebnissen aus Eichhorn (2016) verglichen. Zum Vergleich ist nur Bedingung 5PS geeignet, da in Eichhorn (2016) ausschließlich fünfstufige Likert-Skalen verwendet wurden. Zur Sicherstellung größtmöglicher Vergleichbarkeit wurden aus Eichhorn (2016) nur die Ergebnisse der Bedingung KV (egozentrische und allozentrische Formulierungsperspektiven) herangezogen. Da hier den Probanden keine Balkendiagramme dargeboten wurden, entspricht diese Bedingung der Darstellung in Studie II. Der Vergleich fand auf Basis der aus den Proportional Odds Regressionen ermittelten graphischen Darstellungen statt. Überprüft wurde, wie sich Schwellenabstände sowie Lage von Mittelwert und Median zwischen den Bedingungen unterscheiden.

Zur Überprüfung der parallelen Regressionsannahme (Grundannahmen des Proportional Odds Modells) wurden für alle Merkmale Brant-Tests (Brant, 1990) durchgeführt. Die Annahme besagt, dass die Beziehung zwischen allen ordinalen Stufenpaaren statistisch gleich ist. Dadurch lässt sich der Einfluss der unabhängigen Variable durch einen universellen, für jeden Stufenwechsel gültigen  $\beta$ -Koeffizienten beschreiben (sogenannte *equal slopes*), der von den Abständen zwischen den Stufen unbeeinflusst bleibt. Ein nicht signifikanter Brant-Test spricht für eine Modellgültigkeit (Agresti, 2010; Harrell, 2015). Es handelt sich dabei ebenfalls um ergänzende Analysen, die über die präregistrierten Analysen hinaus gehen.

Für alle Merkmale wurde überprüft, ob das Bildungsniveau ein die Selbstbewertung bestimmender Faktor ist. Die Kategorien *Keine Allgemeine Hochschulreife* und *Allgemeine Hochschulreife* blieben inhaltlich unverändert, wurden aber in *Kein Abitur* bzw. *Abitur* umbenannt. Die Kategorie *Abitur* wurde als Referenzgruppe kodiert. Mithilfe von Proportional Odds Regressionen wurde für alle Merkmale Alter, Körpergewicht, Körpergröße und Einkommen untersucht, ob sich die Stärke des Zusammenhangs zwischen den realen Werten und der Antwortskala in Abhängigkeit des Bildungsniveaus signifikant verändert. Analog zu Studie I wurden die Analysen für ein flexibles Modell ohne äquidistante Schwellen und mit auf den Stichprobenmittelwert zentrierten Werten durchgeführt.

Das Signifikanzniveau wurde bei allen Tests auf  $\alpha = 5\%$  festgelegt. Auf eine Adjustierung des  $\alpha$ -Fehlers wurde verzichtet, da eine Interpretation des Gesamtbildes der Daten mithilfe



eines explorativen Vorgehens im Fokus der Analysen stand und keine klassische Hypothesenprüfung. Alle geschätzten Werte bzw. Parameter waren signifikant von Null verschieden, falls nicht anders angegeben. Alle Signifikanztests wurden zweiseitig durchgeführt.

### 3.3 Ergebnisse Studie II

#### 3.3.1 Deskriptive Statistik

Tabelle 3.3.1. zeigt jeweils für Frauen und Männer Mittelwert, Standardabweichung, Median, kleinsten Wert und größten Wert der realen Werte für alle Merkmale. Dabei ist ersichtlich, dass die Lagemaße der zentralen Tendenz beim Merkmal Einkommen und bei den körperbezogenen Merkmalen bei Frauen geringer ausfallen als bei Männern. Beim Merkmal Alter stimmen diese für beide Geschlechter in etwa überein. Abgesehen von Standardabweichung und größtem Wert beim Merkmal Einkommen sowie kleinste und größte Werte bei den Merkmalen Körpergröße und Körpergewicht fallen die Streuungsmaße für beide Geschlechter annähernd gleich aus.

Tabelle 3.3.1.

*Mittelwert (M), Standardabweichung (SD), Median (Md), Minimum (Min) und Maximum (Max) reale Werte getrennt für Frauen und Männer*

Merkmal	Frauen					Männer				
	<i>M</i>	<i>SD</i>	<i>Md</i>	<i>Min</i>	<i>Max</i>	<i>M</i>	<i>SD</i>	<i>Md</i>	<i>Min</i>	<i>Max</i>
Alter	31.8	10.7	29	18	77	32.8	11.3	29.5	18	75
Einkommen	1 353	1 180	1 200	0	20 000	2 120	3 206	1 690	0	50 000
Körpergröße	168	6.5	168	150	192	181	7.5	181	160	210
Körpergewicht	69.7	17.3	65	38	170	83.3	16.8	80.5	50	160

*Anmerkungen.*

*M* = Mittelwert; *SD* = Standardabweichung; *Md* = Median; *Min* = kleinster Wert; *Max* = größter Wert.

Die Histogramme der Häufigkeiten der realen Werte für die vier erhobenen Merkmale spiegeln die deskriptiven Maße aus Tabelle 3.3.1. wider. Die Merkmale Körpergewicht, Körpergröße und Einkommen weisen bimodale Verteilungen auf. Beim Merkmal Einkommen ist diese weniger eindeutig, da sich deutlich mehr weibliche als männliche

Probanden beteiligten. Beim Merkmal Alter zeigt sich eine unimodale linkssteile Verteilung. Die Histogramme der Häufigkeiten der realen Werte aller vier erhobenen Merkmale finden sich in Anhang B (Abbildungen B.8. und B.9.).

Die Tabellen 3.3.2., 3.3.3. und 3.3.4. zeigen Mittelwert, Standardabweichung und Median für alle Merkmale und Bedingungen getrennt nach Bildungsniveau und Geschlecht. Dabei ist bei beiden Geschlechtern ersichtlich, dass die Lagemaße der zentralen Tendenz beim Merkmal Alter unter allen Bedingungen beim Bildungsniveau *Abitur* insgesamt am geringsten ausfallen und beim Merkmal Körpergewicht unter allen Bedingungen beim Bildungsniveau *Kein Abitur* am höchsten ausfallen. Beim Merkmal Einkommen ergeben sich für Frauen unter allen Bedingungen beim Bildungsniveau *Abitur* die niedrigsten Lagemaße der zentralen Tendenz. Dieses Muster ist bei Männern bis auf zwei Ausnahmen ebenfalls zu beobachten (höhere Mittelwerte unter den Bedingungen 5PS und 2PS bei *Abitur* verglichen zu *Kein Abitur*). Bei Frauen fallen beim Merkmal Einkommen Mittelwert und Median immer geringer aus als bei Männern. Beim Merkmal Körpergröße zeigen sich keine auffälligen Unterschiede in den Lagemaßen der zentralen Tendenz zwischen den Bedingungen.

Tabelle 3.3.2.

Mittelwert (*M*), Standardabweichung (*SD*) und Median (*Md*) der realen Werte getrennt nach Bildungsniveau und Geschlecht – Bedingung 5PS<sup>1</sup>

Merkmal	Kein Abitur				Abitur			
	<i>n</i>	<i>M</i>	<i>SD</i>	<i>Md</i>	<i>n</i>	<i>M</i>	<i>SD</i>	<i>Md</i>
Alter								
Frauen	159	35	12.2	32	299	29	9.8	26
Männer	42	37	14.2	35	89	31	10.2	29
Einkommen								
Frauen	159	1 377	848	1 300	299	1 251	1 327	950
Männer	42	2 140	2 283	1 880	89	2 424	5 577	1 200
Körpergröße								
Frauen	159	167	6.2	167	299	168	6.5	168
Männer	42	180	6.9	180	89	181	7.3	181
Körpergewicht								
Frauen	159	73	17.7	70	299	69	16.9	65
Männer	42	86	21.8	81	89	80	13.0	80

Anmerkungen.

*n* = Anzahl; *M* = Mittelwert; *SD* = Standardabweichung; *Md* = Median.

<sup>1</sup> Fünfstufige Antwortskala.

Tabelle 3.3.3.

Mittelwert (*M*), Standardabweichung (*SD*) und Median (*Md*) der realen Werte getrennt nach Bildungsniveau und Geschlecht – Bedingung 4PS<sup>1</sup>

Merkmal	Kein Abitur				Abitur			
	<i>n</i>	<i>M</i>	<i>SD</i>	<i>Md</i>	<i>n</i>	<i>M</i>	<i>SD</i>	<i>Md</i>
Alter								
Frauen	155	33	9.6	31	326	31	10.2	28
Männer	35	37	12.0	35	78	32	10.5	29
Einkommen								
Frauen	155	1 462	825	1 330	326	1 317	1 084	1 085
Männer	35	2 059	1 021	2 100	78	1 885	1 484	1 850
Körpergröße								
Frauen	155	168	6.3	168	326	168	6.4	168
Männer	35	182	9.3	180	78	181	7.9	180
Körpergewicht								
Frauen	155	74	18.0	69	326	67	13.9	63
Männer	35	89	19.0	85	78	83	14.5	81

Anmerkungen.

*n* = Anzahl; *M* = Mittelwert; *SD* = Standardabweichung; *Md* = Median.

<sup>1</sup> Vierstufige Antwortskala.

Tabelle 3.3.4.

Mittelwert (*M*), Standardabweichung (*SD*) und Median (*Md*) der realen Werte getrennt nach Bildungsniveau und Geschlecht – Bedingung 2PS<sup>1</sup>

Merkmal	Kein Abitur				Abitur			
	<i>n</i>	<i>M</i>	<i>SD</i>	<i>Md</i>	<i>n</i>	<i>M</i>	<i>SD</i>	<i>Md</i>
Alter								
Frauen	164	36	12.6	33	302	30	9.7	28
Männer	45	32	9.3	30	75	32	11.7	27
Einkommen								
Frauen	164	1 485	899	1 418	302	1 352	1 512	1 051
Männer	45	1 812	1 093	1 900	75	2 206	2 610	1 250
Körpergröße								
Frauen	164	168	6.4	168	302	168	6.7	168
Männer	45	183	7.5	183	75	181	6.8	182
Körpergewicht								
Frauen	164	74	20.6	69	302	68	17.3	63
Männer	45	88	20.9	85	75	80	15.1	77

Anmerkungen.

*n* = Anzahl; *M* = Mittelwert; *SD* = Standardabweichung; *Md* = Median.

<sup>1</sup> Dichotome Antwortskala.

Die Tabellen 3.3.5. und 3.3.6. zeigen Mittelwert, Standardabweichung und Median der beobachteten Werte für alle Merkmale und Bedingungen getrennt nach Geschlecht. Dabei ist ersichtlich, dass beim Merkmal Alter bei Frauen Mittelwert und Median unter Bedingung 5PS bei allozentrischer Perspektive im Vergleich zur egozentrischen Perspektive geringer ausfallen und die Standardabweichung höher. Unter den Bedingungen 4PS und 2PS fallen die Mittelwerte bei allozentrischer Perspektive im Vergleich zur egozentrischen Perspektive ebenfalls geringer aus. Die Standardabweichung fällt unter Bedingung 4PS bei allozentrischer Perspektive auffallend höher aus (*SD*: 1.72 vs. 0.58). Bei Männern zeigt sich unter Bedingung 5PS ein ähnliches Muster wie bei Frauen. Unter Bedingung 2PS ist bei Männern im Vergleich zu Frauen ein gegenteiliges Muster zu beobachten. Unter Bedingung 4PS führt der Perspektivenwechsel nicht zu auffallend divergierenden Standardabweichungen und Maßen der zentralen Tendenz.

Beim Merkmal Einkommen fallen Mittelwert, Median und Standardabweichung bei Männern unter Bedingung 5PS bei allozentrischer Perspektive im Vergleich zur

egozentrischen Perspektive höher aus. Unter Bedingung 4PS fällt dabei der Mittelwert ebenfalls höher aus. Unter Bedingung 2PS führt der Perspektivenwechsel nicht zu auffälligen Unterschieden bei den Maßen der zentralen Tendenz und den Standardabweichungen. Bei Frauen zeigt sich ein ähnliches, allerdings weniger stark ausgeprägtes Muster wie bei Männern.

Beim Merkmal Körpergröße zeigen sich innerhalb beider Geschlechter keine auffälligen Unterschiede. Unter keiner Bedingung führt der Perspektivenwechsel zu auffallend divergierenden Standardabweichungen und Maßen der zentralen Tendenz. Beim Merkmal Körpergewicht fallen Mittelwert und Median bei Frauen unter allen Bedingungen bei allozentrischer Perspektive im Vergleich zur egozentrischen Perspektive deutlich geringer aus. Bei Männern zeigt sich ein ähnliches, allerdings weniger stark ausgeprägtes Muster wie bei Frauen.

Tabelle 3.3.5.

*Mittelwert (M), Standardabweichung (SD) und Median (Md) beobachtete Werte für Frauen*

Merkmal und Item-Nr.		5PS <sup>1</sup>			4PS <sup>2</sup>			2PS <sup>3</sup>		
		M	SD	Md	M	SD	Md	M	SD	Md
Alter										
	1(a)	2.58	0.76	3.00	2.18	0.58	2.00	0.25	0.44	0.00
	1(b)	2.29	0.90	2.00	1.94	1.72	2.00	0.15	0.36	0.00
Einkommen										
	2(a)	2.08	1.08	2.00	1.77	0.82	2.00	0.21	0.41	0.00
	2(b)	2.20	1.17	2.00	1.92	0.98	2.00	0.29	0.46	0.00
Körpergröße										
	3(a)	2.88	1.01	3.00	2.38	0.84	2.00	0.45	0.50	0.00
	3(b)	2.85	1.06	3.00	2.36	0.88	2.00	0.44	0.50	0.00
Körpergewicht										
	4(a)	3.60	1.09	4.00	2.89	0.89	3.00	0.68	0.47	1.00
	4(b)	3.00	1.10	3.00	2.37	0.91	2.00	0.42	0.94	0.00

*Anmerkungen.*

*M* = Mittelwert; *SD* = Standardabweichung; *Md* = Median.

<sup>1</sup> Fünfstufige Antwortskala.

<sup>2</sup> Vierstufige Antwortskala.

<sup>3</sup> Dichotome Antwortskala.

(a) Egozentrische Perspektive; (b) Allozentrische Perspektive.

Tabelle 3.3.6.

*Mittelwert (M), Standardabweichung (SD) und Median (Md) beobachtete Werte für Männer*

Merkmal und Item-Nr.		5PS <sup>1</sup>			4PS <sup>2</sup>			2PS <sup>3</sup>		
		M	SD	Md	M	SD	Md	M	SD	Md
Alter										
	1(a)	2.76	0.79	3.00	2.13	0.65	2.00	0.19	0.40	0.00
	1(b)	2.61	0.93	3.00	2.11	0.70	2.00	0.28	0.45	0.00
Einkommen										
	2(a)	2.21	1.21	2.00	2.00	0.93	2.00	0.32	0.47	0.00
	2(b)	2.56	1.34	3.00	2.19	0.99	2.00	0.39	0.49	0.00
Körpergröße										
	3(a)	3.10	0.93	3.00	2.67	0.76	3.00	0.60	0.49	1.00
	3(b)	3.19	1.00	3.00	2.64	0.84	3.00	0.62	0.49	1.00
Körpergewicht										
	4(a)	3.37	1.12	3.00	2.65	0.84	3.00	0.54	0.50	1.00
	4(b)	2.89	1.13	3.00	2.35	0.83	2.00	0.34	0.48	0.00

*Anmerkungen.**M* = Mittelwert; *SD* = Standardabweichung; *Md* = Median.<sup>1</sup> Fünfstufige Antwortskala.<sup>2</sup> Vierstufige Antwortskala.<sup>3</sup> Dichotome Antwortskala.

(a) Egozentrische Perspektive; (b) Allozentrische Perspektive.

Die Tabelle 3.3.7. zeigt die Ergebnisse der Spearman- bzw. punktbiserialen Korrelationen (Beziehung zwischen den gewählten Antwortkategorien und den realen Werten der jeweiligen Substichprobe) für alle Merkmale und Bedingungen getrennt nach Geschlecht. Beim Merkmal Alter fallen im Vergleich zu allen anderen Merkmalen die Korrelationen bei beiden Geschlechtern deutlich geringer aus. Bei Männern ergeben sich unter den Bedingungen 4PS (allozentrische Perspektive) und 2PS (egozentrische und allozentrische Perspektive) Korrelationskoeffizienten, die sich von Null nicht signifikant unterscheiden. Beim Merkmal Einkommen fallen bei Frauen die Korrelationen unter Bedingung 2PS geringer aus im Vergleich zu den Bedingungen 5PS und 4PS. Bei den Merkmalen Einkommen, Körpergröße und Körpergewicht ergeben sich bei Männern zwischen allen Bedingungen keine systematischen Unterschiede bei den Korrelationskoeffizienten. Bei den Merkmalen Körpergröße und Körpergewicht zeigt sich bei Frauen ein ähnliches Muster wie bei Männern.

Tabelle 3.3.7.

Ergebnisse Spearman- bzw. punktbiseriale Korrelationen der Bedingungen 5PS<sup>1</sup>, 4PS<sup>2</sup> und 2PS<sup>3</sup> für Frauen und Männer

Merkmal und Item-Nr.	Frauen						Männer					
	5PS		4PS		2PS		5PS		4PS		2PS	
	$r_s$	$p$	$r_s$	$p$	$r_{pb}$	$p$	$r_s$	$p$	$r_s$	$p$	$r_{pb}$	$p$
Alter												
1(a)	.28	< .001	.24	< .001	.28	< .001	.47	< .001	.41	< .001	.08	.36
1(b)	.22	< .001	.15	.001	.17	< .001	.24	.005	.04	.64	-.003	.97
Einkommen												
2(a)	.54	< .001	.61	< .001	.40	< .001	.70	< .001	.65	< .001	.57	< .001
2(b)	.59	< .001	.64	< .001	.45	< .001	.69	< .001	.69	< .001	.60	< .001
Körpergröße												
3(a)	.76	< .001	.72	< .001	.74	< .001	.73	< .001	.73	< .001	.73	< .001
3(b)	.79	< .001	.74	< .001	.76	< .001	.74	< .001	.70	< .001	.71	< .001
Körpergewicht												
4(a)	.71	< .001	.68	< .001	.60	< .001	.67	< .001	.70	< .001	.69	< .001
4(b)	.73	< .001	.71	< .001	.71	< .001	.63	< .001	.55	< .001	.63	< .001

Anmerkungen.

$r_s$  = Spearmans Rangkorrelationskoeffizient;  $r_{pb}$  = punktbiserialer Korrelationskoeffizient (Zusammenhang zwischen gewählten Antwortkategorien und realen Werten).

<sup>1</sup> Fünfstufige Antwortskala.

<sup>2</sup> Vierstufige Antwortskala.

<sup>3</sup> Dichotome Antwortskala.

Signifikante Ergebnisse ( $p \leq .05$ ) sind fett gedruckt.

(a) Egozentrische Perspektive; (b) Allozentrische Perspektive.

### 3.3.2 Regressionsanalyse

In den Abbildungen 3.3.1. bis 3.3.8. ist für jede Bedingung die Lage der aus den Proportional Odds Regressionen ermittelten Schwellenparameter (5PS und 4PS) bzw. der aus den logistischen Regressionen ermittelten logistischen Funktionen (2PS) graphisch (rot) veranschaulicht. In den Abbildungen 3.3.9. bis 3.3.12. wird Bedingung 5PS den Ergebnissen aus Eichhorn (2016) gegenübergestellt. In jedem Diagramm sind Mittelwert (blau) und Median (grün) der realen Werte für die jeweilige Teilstichprobe des nicht-messwiederholten Faktors abgebildet. In Fällen, in denen der Mittelwert und/oder der Median auf einer Schwelle liegen, konnte aus technischen Gründen nur der Mittelwert bzw. der Median angezeigt werden. Wenn Mittelwert und Median aufeinanderliegen, konnte nur der Median

abgebildet werden. Die Tabellen 3.3.8. und 3.3.9. berichten die Schwellenparameter und Steigungskoeffizienten für jede Bedingung (5PS und 4PS). Die Tabelle 3.3.10. berichtet die Ergebnisse der logistischen Regressionen (2PS). Die Ergebnisse der Likelihood Ratio Tests, mit denen die Annahme äquidistanter Schwellen überprüft wurde, sind in Anhang B einzusehen (Tabelle B.2.). Die Tabellen 3.3.11., 3.3.12. und 3.3.13. zeigen für die Bedingungen 5PS, 4PS und 2PS und für alle untersuchten Merkmale die Ergebnisse der Proportional Odds Regressionen bzw. logistischen Regressionen mit Bildungsniveau als zusätzlichen Prädiktor für die Stärke des Zusammenhangs zwischen den realen Werten und der Antwortskala.

*Alter.* Die Abbildungen 3.3.1. und 3.3.2. zeigen, dass unter Bedingung 5PS Median und Mittelwert nur bei Item 1(a) bei Männern eindeutig in der Mittelkategorie liegen. Nur bei Männern weisen beide Formulierungsvarianten eine Annäherung an äquidistante Schwellen auf, welche sich in nicht signifikanten Likelihood Ratio Tests widerspiegeln. Bei beiden Geschlechtern zeigen sich zwischen egozentrischer und allozentrischer Fragenformulierung keine auffallend systematischen Unterschiede. Unter Bedingung 4PS korrespondieren bei beiden Geschlechtern bei beiden Formulierungsvarianten Median und Mittelwert nicht mit der zweiten Schwelle. Bei Item 1(a) konnten bei Männern nur zwei Schwellen ermittelt werden, da in dieser Teilstichprobe die Antwortkategorie 4 (*hoch*) unbesetzt ist. Aus diesem Grund ergibt sich beim Likelihood Ratio Test ein  $p$ -Wert von 1, da zwei Schwellen immer einen konstanten Abstand aufweisen. Unter Bedingung 2PS erweist sich nur bei Frauen das Modell der logistischen Regression als angemessen. Dabei korrespondieren Median und Mittelwert nicht mit dem Wendepunkt der logistischen Funktion. Bei Männern besteht kein signifikanter Zusammenhang zwischen Alter und dichotomer Antwortskala.

*Einkommen.* Die Abbildungen 3.3.3. und 3.3.4. zeigen, dass unter Bedingung 5PS bei beiden Geschlechtern Median und Mittelwert nicht in der mittleren Antwortkategorie liegen (außer Mittelwert bei Item 2(b) bei Männern). Nur bei Männern weist die egozentrische Fragenformulierung eine Annäherung an äquidistante Schwellen auf. Der entsprechende Likelihood Ratio Test ist nicht signifikant. Bei Frauen zeigt der Perspektivenwechsel durch die Itemformulierung keinen Einfluss. Unter Bedingung 4PS korrespondieren bei beiden Formulierungsvarianten bei beiden Geschlechtern Median und Mittelwert nicht mit der zweiten Schwelle. Bei Frauen liegen bei beiden Formulierungsvarianten Median und



Mittelwert nahe der ersten Schwelle, bei Männern tendenziell nahe der zweiten Schwelle. Für beide Geschlechter ergibt sich bei egozentrischer Fragenformulierung und bei Männern zusätzlich bei allozentrischer Fragenformulierung eine Annäherung an äquidistante Schwellen, bestätigt durch nicht signifikante Likelihood Ratio Tests. Median und Mittelwert korrespondieren bei Frauen unter Bedingung 2PS bei beiden Formulierungsvarianten nicht mit dem Wendepunkt der logistischen Funktion. Bei Männern korrespondieren Median und Mittelwert nur bei allozentrischer Formulierung mit dem Wendepunkt der logistischen Funktion.

*Körpergröße.* Die Abbildungen 3.3.5. und 3.3.6. zeigen, dass unter Bedingung 5PS Median und Mittelwert bei beiden Formulierungsvarianten bei beiden Geschlechtern in der mittleren Antwortkategorie liegen. Bei Männern weisen beide Formulierungsvarianten eine Annäherung an äquidistante Schwellen auf, bestätigt durch nicht signifikante Likelihood Ratio Tests. Unter Bedingung 4PS liegen Median und Mittelwert bei beiden Formulierungsvarianten bei beiden Geschlechtern nahe der zweiten Schwelle. Beide Formulierungsvarianten weisen bei beiden Geschlechtern eine Annäherung an äquidistante Schwellen auf. Die entsprechenden Likelihood Ratio Tests sind nicht signifikant. Unter Bedingung 2PS korrespondieren Median und Mittelwert bei beiden Geschlechtern bei beiden Formulierungsvarianten mit dem Wendepunkt der logistischen Funktion. Unter keiner Bedingung hat der Perspektivenwechsel einen Einfluss.

*Körpergewicht.* Die Abbildungen 3.3.7. und 3.3.8. zeigen, dass unter Bedingung 5PS bei Männern Median und Mittelwert bei beiden Formulierungsvarianten in der mittleren Antwortkategorie liegen. Bei Frauen ist dies nur bei allozentrischer Formulierung der Fall. Unter Bedingung 4PS korrespondieren Median und Mittelwert bei beiden Formulierungsvarianten bei Männern annähernd mit der zweiten Schwelle. Bei Frauen ist dies nur bei allozentrischer Formulierung zu beobachten. Median und Mittelwert liegen bei beiden Geschlechtern bei egozentrischer Formulierung über der zweiten Schwelle und bei allozentrischer Formulierung unter der zweiten Schwelle. Bei den Bedingungen 5PS und 4PS ergibt sich bei beiden Formulierungsvarianten bei beiden Geschlechtern eine Annäherung an äquidistante Schwellen, bestätigt durch nicht signifikante Likelihood Ratio Tests (außer Item 4(b) bei Frauen). Unter Bedingung 2PS korrespondieren Median und Mittelwert bei beiden Geschlechtern bei beiden Formulierungsvarianten mit dem Wendepunkt der

logistischen Funktion. Bei Frauen liegt darüber hinaus bei allozentrischer Formulierung der Mittelwert auf dem Wendepunkt der logistischen Funktion.

*Brant-Test.* Bei Männern ergaben sich bei 12 von 15 Brant-Tests nicht signifikante Ergebnisse (signifikante Ergebnisse lagen ausschließlich beim Merkmal Einkommen vor). Bei Frauen ergaben sich über alle Merkmale überwiegend signifikante Ergebnisse (6 von 16 Brant-Tests nicht signifikant). Die Ergebnisse aller Brant-Tests sind in Anhang B einzusehen (Tabelle B.3.).

*Bildungsniveau.* Bei den Merkmalen Einkommen, Körpergröße und Körpergewicht (beide Geschlechter) und beim Merkmal Alter (Frauen) zeigt sich ein Zusammenhang zwischen den realen Werten und der Antwortskala (signifikante Effekte für den Prädiktor *reale Werte* in der Referenzgruppe *Abitur* unter allen Bedingungen). Beim Merkmal Alter ergibt sich bei Männern unter den Bedingungen 2PS (egozentrische und allozentrische Fragenformulierung) und 4PS (allozentrische Fragenformulierung) kein Zusammenhang zwischen den realen Werten und der Antwortskala (keine signifikanten Effekte für den Prädiktor *reale Werte* in der Referenzgruppe *Abitur*). Bei allen Merkmalen unter allen Bedingungen gibt es bei beiden Geschlechtern keine Anzeichen dafür, dass sich die Stärke des Zusammenhangs in Abhängigkeit des Bildungsniveaus ändert.

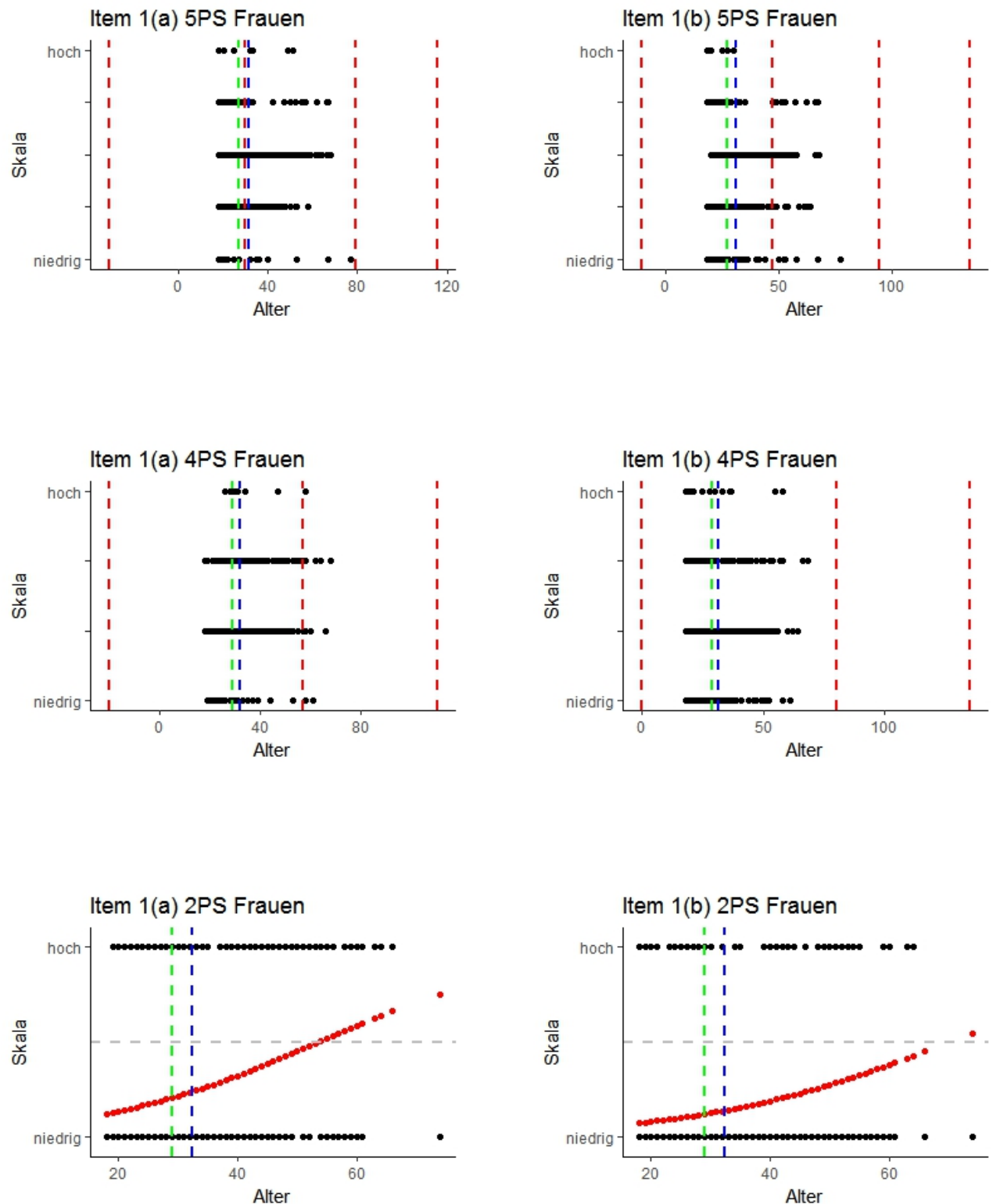


Abbildung 3.3.1. Regression Antwortskala und reale Werte für das Merkmal Alter für Frauen.

Alter in Jahren; schwarze Punkte = gewählte Antwortkategorien; rote Linien = geschätzte Schwellen zwischen den Kategorien der Skala; rote Punkte = vorhergesagte Werte auf der dichotomen Skala für jeden empirischen Wert (realen Wert); blaue Linie = Mittelwert der realen Werte; grüne Linie = Median der realen Werte.

5PS = Fünfstufige Antwortskala.

4PS = Vierstufige Antwortskala.

2PS = Dichotome Antwortskala.

(a) Egozentrische Perspektive; (b) Allozentrische Perspektive.

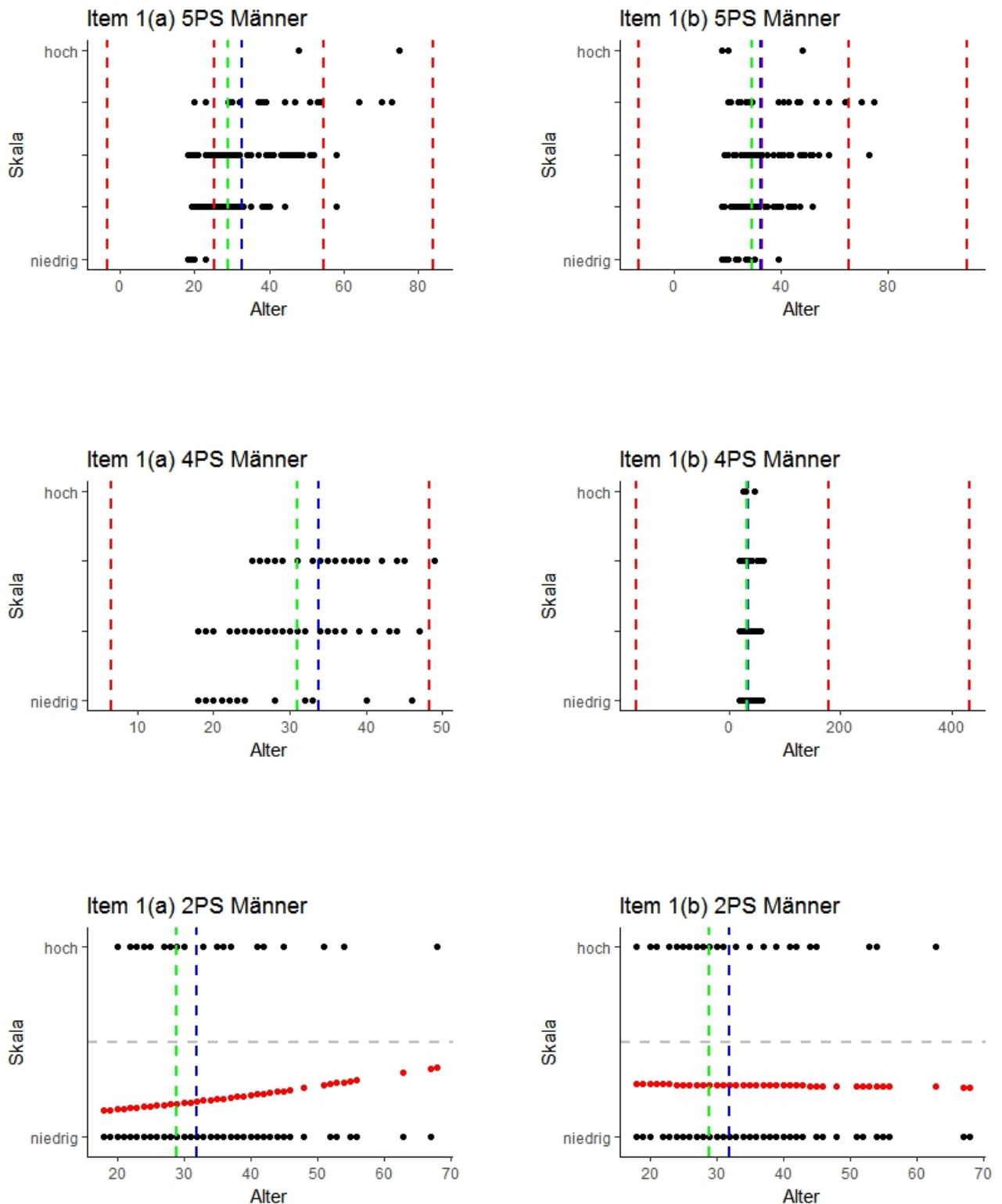


Abbildung 3.3.2. Regression Antwortskala und reale Werte für das Merkmal Alter für Männer.

Alter in Jahren; schwarze Punkte = gewählte Antwortkategorien; rote Linien = geschätzte Schwellen zwischen den Kategorien der Skala; rote Punkte = vorhergesagte Werte auf der dichotomen Skala für jeden empirischen Wert (realen Wert); blaue Linie = Mittelwert der realen Werte; grüne Linie = Median der realen Werte.

5PS = Fünfstufige Antwortskala.

4PS = Vierstufige Antwortskala.

2PS = Dichotome Antwortskala.

(a) Egozentrische Perspektive; (b) Allozentrische Perspektive.

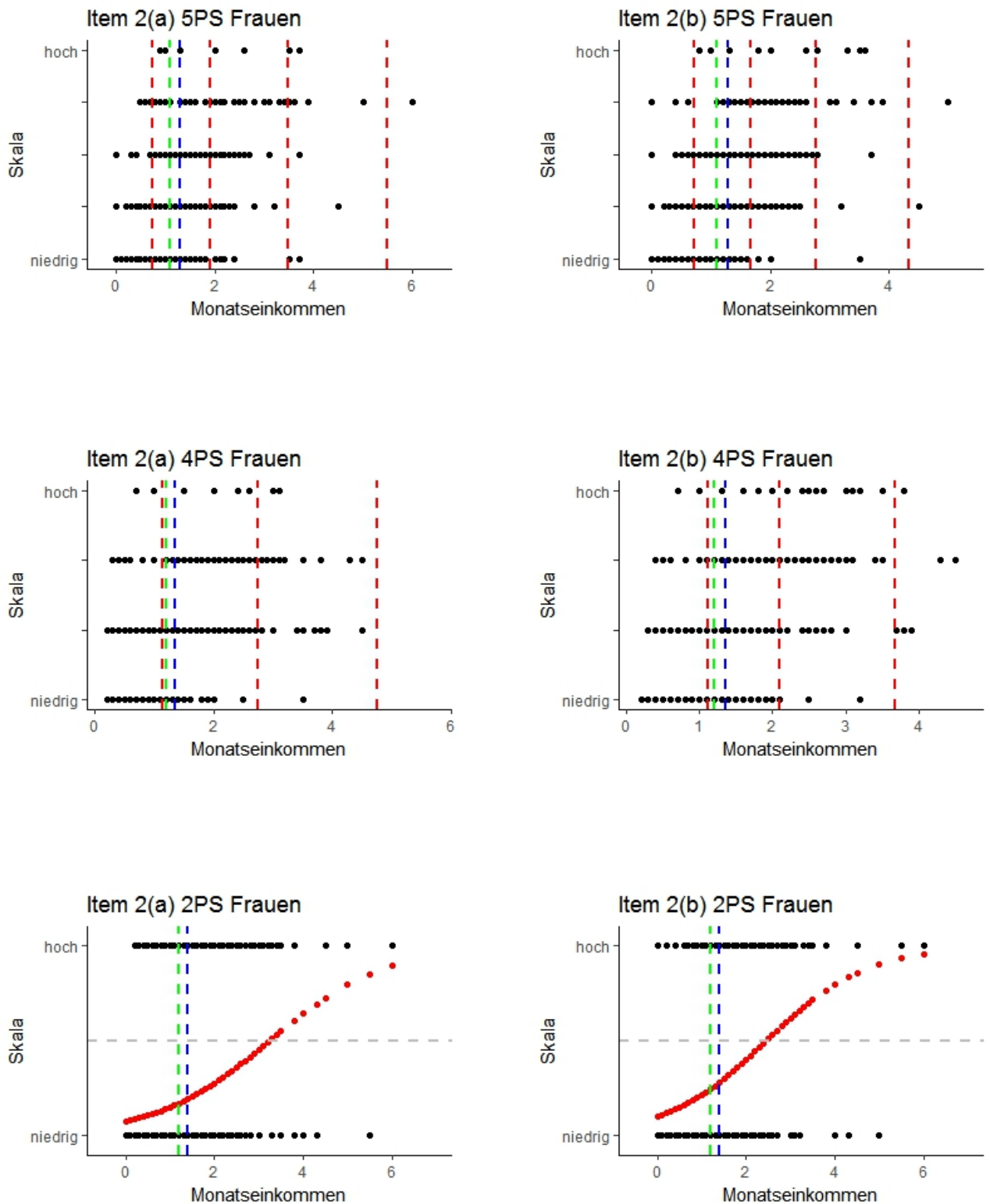


Abbildung 3.3.3. Regression Antwortskala und reale Werte für das Merkmal Einkommen für Frauen. Einkommen in Tausend-Euro-Werten; schwarze Punkte = gewählte Antwortkategorien; rote Linien = geschätzte Schwellen zwischen den Kategorien der Skala; rote Punkte = vorhergesagte Werte auf der dichotomen Skala für jeden empirischen Wert (realen Wert); blaue Linie = Mittelwert der realen Werte; grüne Linie = Median der realen Werte.

5PS = Fünfstufige Antwortskala.

4PS = Vierstufige Antwortskala.

2PS = Dichotome Antwortskala.

(a) Egozentrische Perspektive; (b) Allozentrische Perspektive.

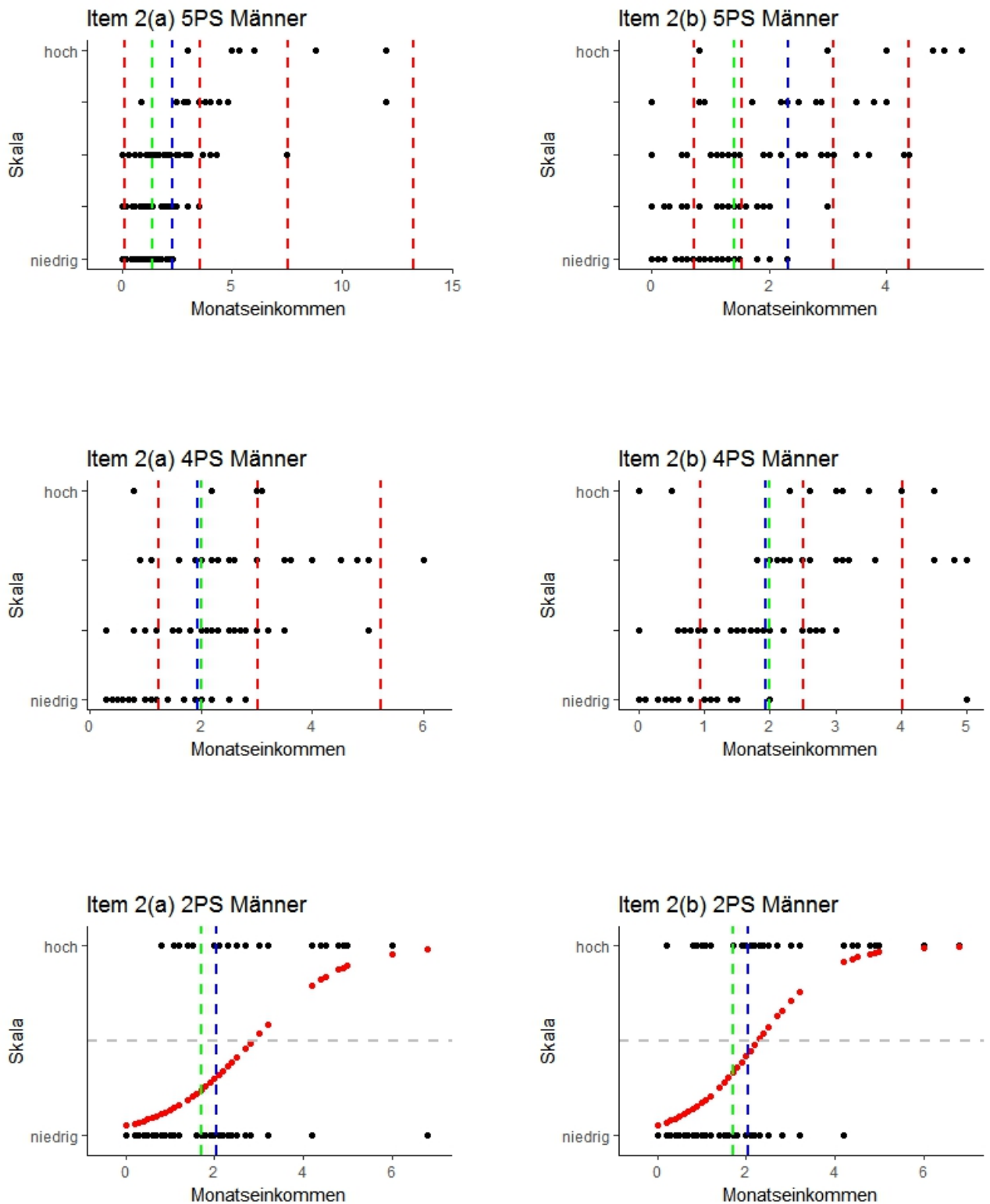
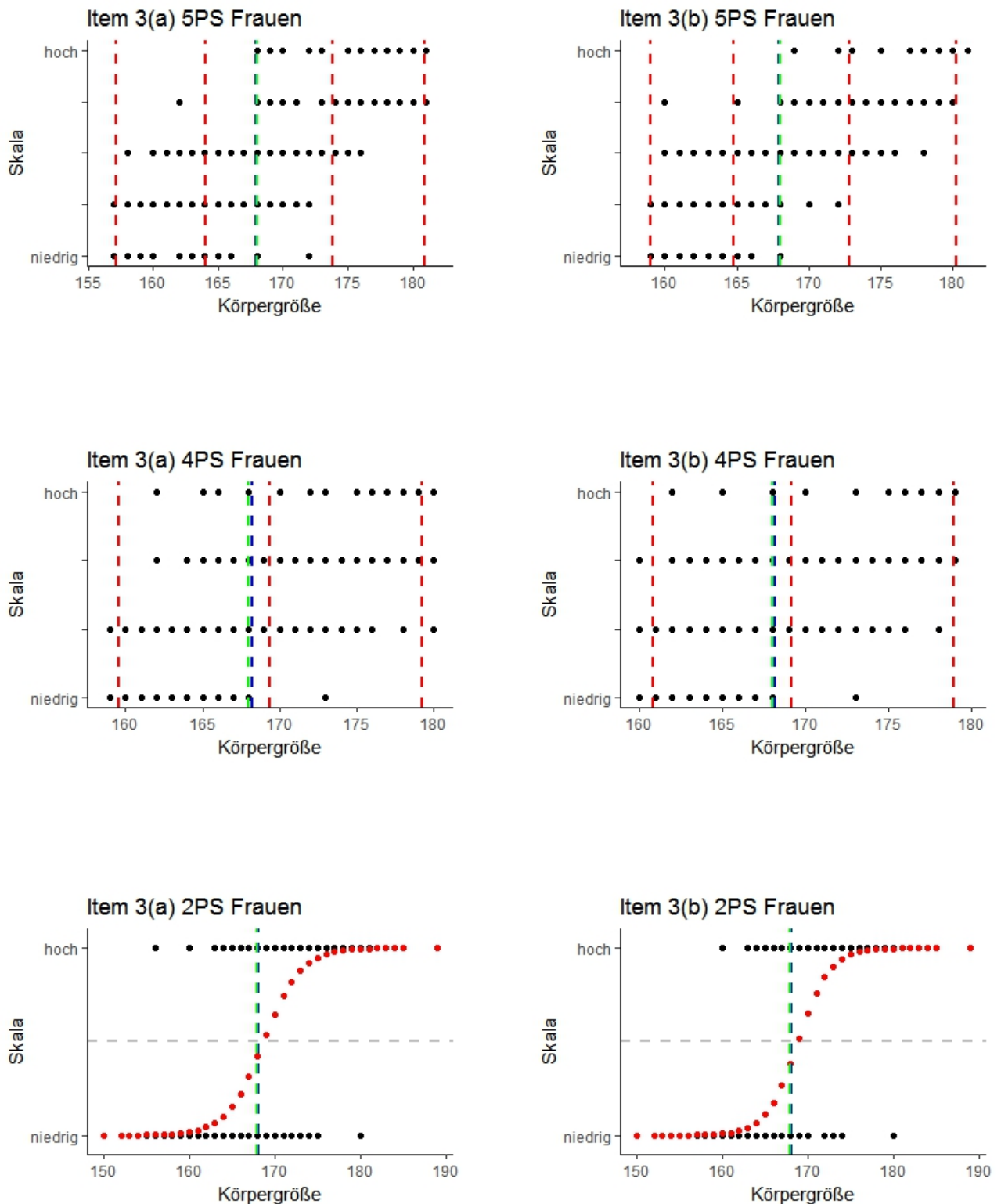
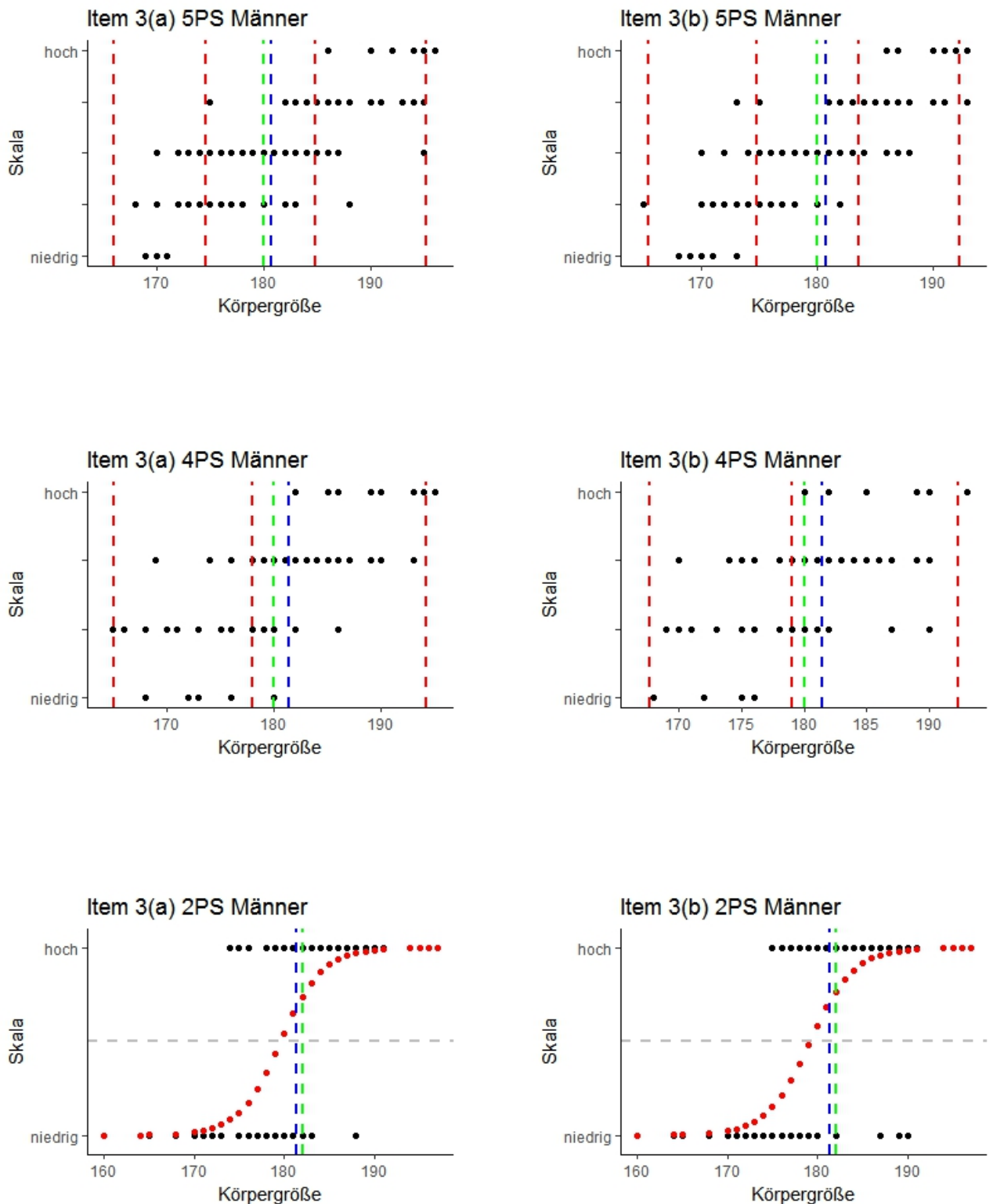


Abbildung 3.3.4. Regression Antwortskala und reale Werte für das Merkmal Einkommen für Männer. Einkommen in Tausend-Euro-Werten; schwarze Punkte = gewählte Antwortkategorien; rote Linien = geschätzte Schwellen zwischen den Kategorien der Skala; rote Punkte = vorhergesagte Werte auf der dichotomen Skala für jeden empirischen Wert (realen Wert); blaue Linie = Mittelwert der realen Werte; grüne Linie = Median der realen Werte. 5PS = Fünfstufige Antwortskala. 4PS = Vierstufige Antwortskala. 2PS = Dichotome Antwortskala. (a) Egozentrische Perspektive; (b) Allozentrische Perspektive.



**Abbildung 3.3.5.** Regression Antwortskala und reale Werte für das Merkmal Körpergröße für Frauen. Körpergröße in Zentimetern; schwarze Punkte = gewählte Antwortkategorien; rote Linien = geschätzte Schwellen zwischen den Kategorien der Skala; rote Punkte = vorhergesagte Werte auf der dichotomen Skala für jeden empirischen Wert (realen Wert); blaue Linie = Mittelwert der realen Werte; grüne Linie = Median der realen Werte. 5PS = Fünfstufige Antwortskala. 4PS = Vierstufige Antwortskala. 2PS = Dichotome Antwortskala. (a) Egozentrische Perspektive; (b) Allozentrische Perspektive.



**Abbildung 3.3.6.** Regression Antwortskala und reale Werte für das Merkmal Körpergröße für Männer. Körpergröße in Zentimetern; schwarze Punkte = gewählte Antwortkategorien; rote Linien = geschätzte Schwellen zwischen den Kategorien der Skala; rote Punkte = vorhergesagte Werte auf der dichotomen Skala für jeden empirischen Wert (realen Wert); blaue Linie = Mittelwert der realen Werte; grüne Linie = Median der realen Werte. 5PS = Fünfstufige Antwortskala. 4PS = Vierstufige Antwortskala. 2PS = Dichotome Antwortskala. (a) Egozentrische Perspektive; (b) Allozentrische Perspektive.



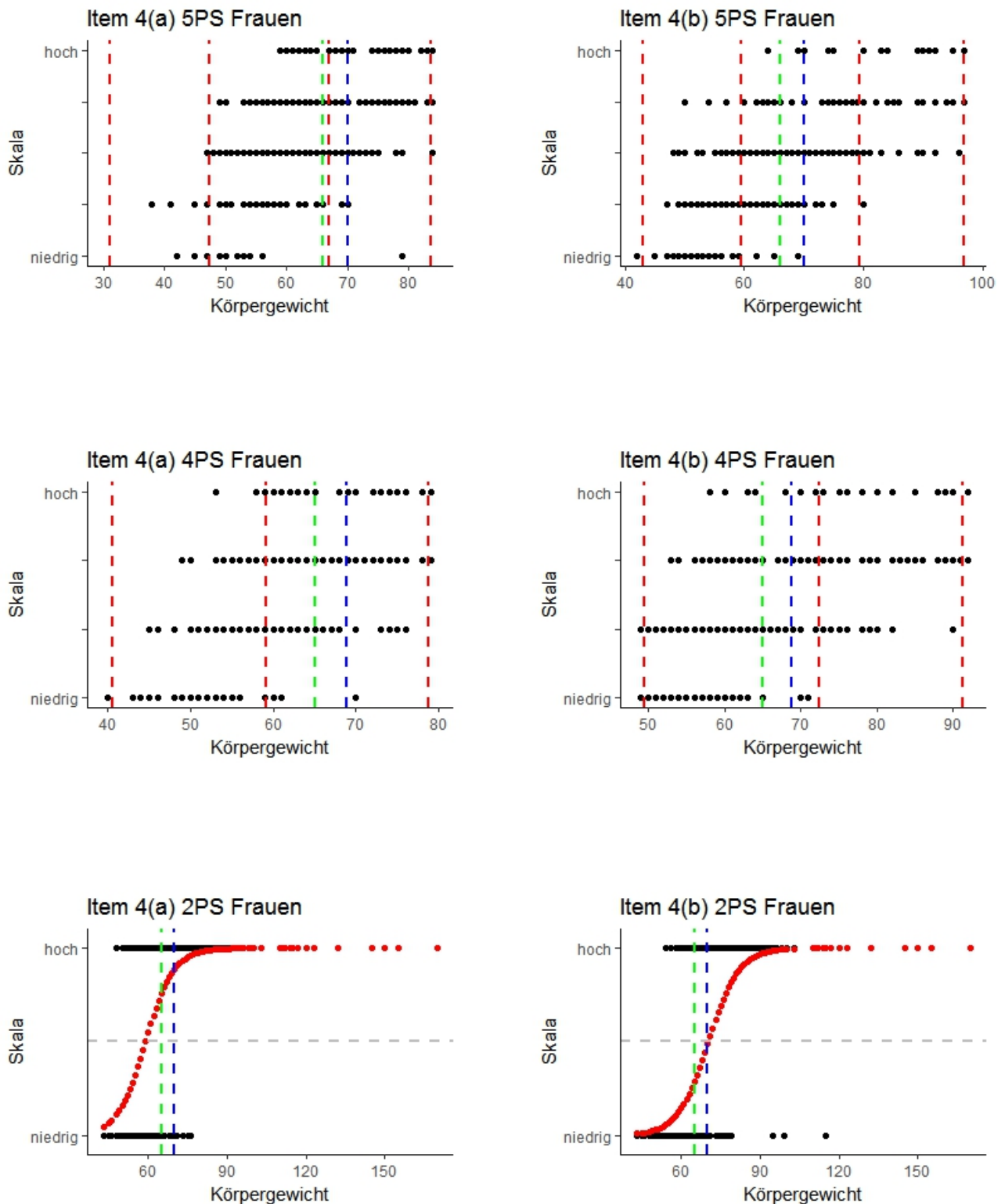
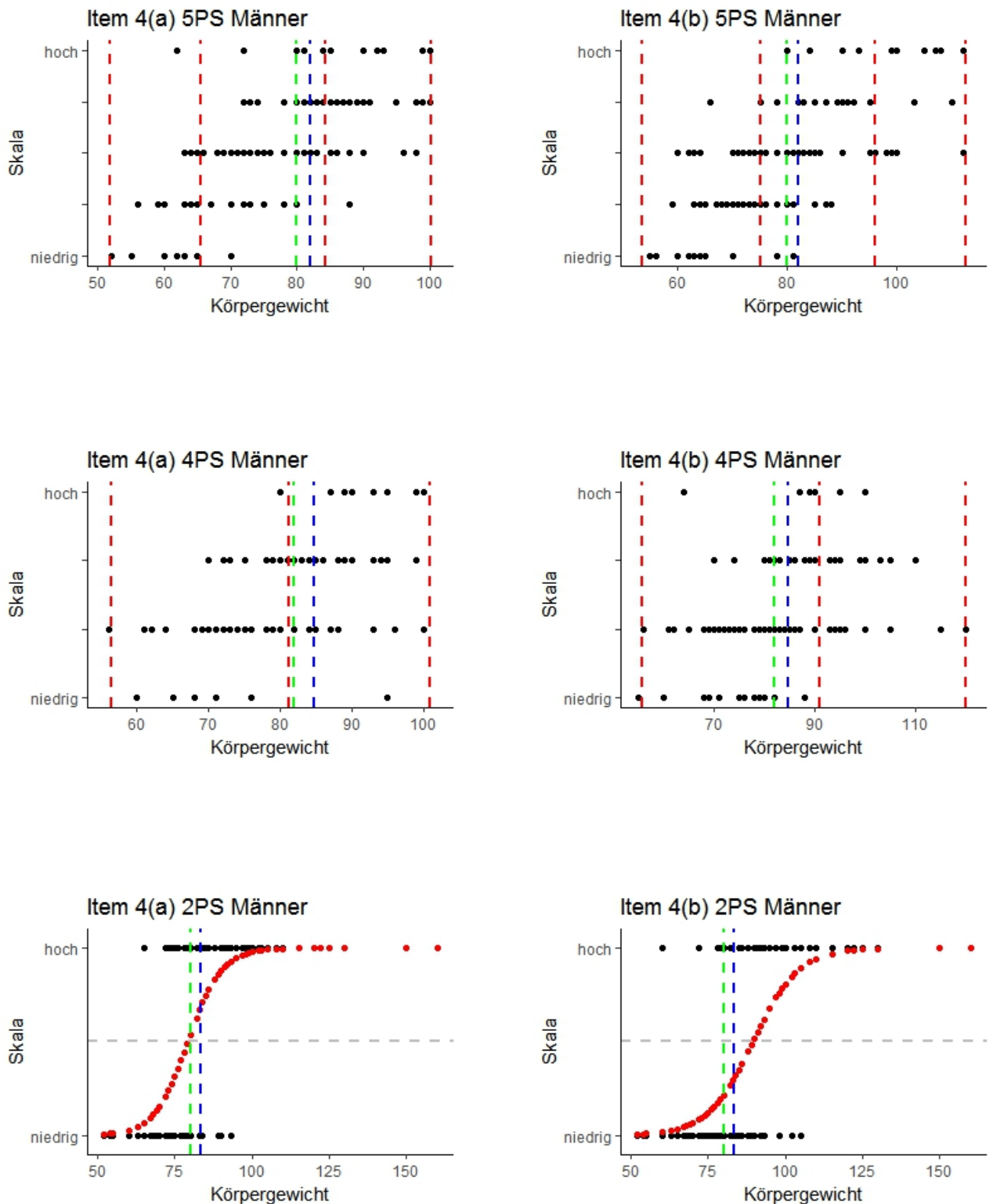


Abbildung 3.3.7. Regression Antwortskala und reale Werte für das Merkmal Körpergewicht für Frauen. Körpergewicht in Kilogramm; schwarze Punkte = gewählte Antwortkategorien; rote Linien = geschätzte Schwellen zwischen den Kategorien der Skala; rote Punkte = vorhergesagte Werte auf der dichotomen Skala für jeden empirischen Wert (realen Wert); blaue Linie = Mittelwert der realen Werte; grüne Linie = Median der realen Werte. 5PS = Fünfstufige Antwortskala. 4PS = Vierstufige Antwortskala. 2PS = Dichotome Antwortskala. (a) Egozentrische Perspektive; (b) Allozentrische Perspektive.



**Abbildung 3.3.8.** Regression Antwortskala und reale Werte für das Merkmal Körpergewicht für Männer. Körpergewicht in Kilogramm; schwarze Punkte = gewählte Antwortkategorien; rote Linien = geschätzte Schwellen zwischen den Kategorien der Skala; rote Punkte = vorhergesagte Werte auf der dichotomen Skala für jeden empirischen Wert (realen Wert); blaue Linie = Mittelwert der realen Werte; grüne Linie = Median der realen Werte. 5PS = Fünfstufige Antwortskala. 4PS = Vierstufige Antwortskala. 2PS = Dichotome Antwortskala. (a) Egozentrische Perspektive; (b) Allozentrische Perspektive.

Tabelle 3.3.8.

*Schwellenparameter ( $t^1$ ) und Steigung ( $s^2$ ) der Bedingung 5PS<sup>3</sup> für Frauen und Männer*

Merkmal und Item-Nr.	Frauen					Männer				
	$t_1$	$t_2$	$t_3$	$t_4$	$s$	$t_1$	$t_2$	$t_3$	$t_4$	$s$
Alter										
1(a)	-1.596	1.538	4.071	5.968	0.052	-0.353	2.560	5.500	8.482	0.101
1(b)	-0.421	1.844	3.666	5.237	0.039	-0.689	1.724	3.432	5.731	0.052
Einkommen										
2(a)	0.851	2.174	3.937	6.208	1.134	0.057	1.197	2.553	4.480	0.338
2(b)	1.009	2.341	3.869	6.058	1.400	0.982	2.108	4.272	6.058	1.377
Körpergröße										
3(a)	54.284	56.646	60.014	62.454	0.345	54.279	57.050	60.401	63.814	0.327
3(b)	65.283	67.650	70.944	73.972	0.410	54.579	57.656	60.575	63.439	0.330
Körpergewicht										
4(a)	3.569	5.457	7.691	9.620	0.115	6.731	8.496	10.937	12.993	0.130
4(b)	5.619	7.776	10.390	12.688	0.131	5.475	7.664	9.798	11.481	0.102

*Anmerkungen.*

<sup>1</sup> Schwellenparameter = Punkt, an dem die Wahrscheinlichkeit für die Wahl der nächsthöheren Antwortkategorie größer wird als für die Wahl der vorherigen Antwortkategorie.

<sup>2</sup> Steigung = Steigungskoeffizient der Regression.

<sup>3</sup> Fünfstufige Antwortskala.

$t_1$  = erste Schwelle;  $t_2$  = zweite Schwelle;  $t_3$  = dritte Schwelle;  $t_4$  = vierte Schwelle.

(a) Egozentrische Perspektive; (b) Allozentrische Perspektive.

Tabelle 3.3.9.

*Schwellenparameter ( $t^1$ ) und Steigung ( $s^2$ ) der Bedingung 4PS<sup>3</sup> für Frauen und Männer*

Merkmal und Item-Nr.	Frauen				Männer			
	$t_1$	$t_2$	$t_3$	$s$	$t_1$	$t_2$	$t_3$	$s$
Alter								
1(a)	-1.027	2.898	5.615	0.051	0.467	3.432	-	0.071
1(b)	6.528 <sup>e-05</sup>	2.725 <sup>e+00</sup>	4.586 <sup>e+00</sup>	0.034	-1.385	1.493	3.593	0.008
Einkommen								
2(a)	1.555	3.725	6.453	1.359	1.421	3.433	5.932	1.136
2(b)	1.710	3.203	5.618	1.529	1.398	3.706	5.901	1.471
Körpergröße								
3(a)	54.473	57.803	61.179	0.341	42.714	46.053	50.251	0.259
3(b)	57.145	60.102	63.567	0.355	45.408	48.507	52.094	0.271
Körpergewicht								
4(a)	5.271	7.682	10.247	0.130	8.163	11.781	14.633	0.145
4(b)	6.420	9.396	11.827	0.130	4.489	7.337	9.667	0.081

*Anmerkungen.*

<sup>1</sup> Schwellenparameter = Punkt, an dem die Wahrscheinlichkeit für die Wahl der nächsthöheren Antwortkategorie größer wird als für die Wahl der vorherigen Antwortkategorie.

<sup>2</sup> Steigung = Steigungskoeffizient der Regression.

<sup>3</sup> Vierstufige Antwortskala.

$t_1$  = erste Schwelle;  $t_2$  = zweite Schwelle;  $t_3$  = dritte Schwelle.

(a) Egozentrische Perspektive; (b) Allozentrische Perspektive.

Tabelle 3.3.10.

*Ergebnisse logistische Regression der Bedingung 2PS<sup>1</sup> für Frauen und Männer*

Merkmal und Item-Nr.	Frauen				Männer			
	$\beta$	SE	z	p	$\beta$	SE	z	p
Alter – 1(a)								
Intercept	-2.928	0.352	-8.31	<b>&lt; .001</b>	-2.283	0.719	-3.18	<b>.001</b>
Alter	0.055	0.010	5.71	<b>&lt; .001</b>	0.026	0.020	1.28	.20
Alter – 1(b)								
Intercept	-3.368	0.413	-8.16	<b>&lt; .001</b>	-0.916	0.641	-1.43	.15
Alter	0.048	0.011	4.47	<b>&lt; .001</b>	-0.002	0.019	-0.09	.93
Einkommen – 2(a)								
Intercept	-2.511	0.243	-10.33	<b>&lt; .001</b>	-2.858	0.503	-5.68	<b>&lt; .001</b>
Einkommen	0.774	0.130	5.97	<b>&lt; .001</b>	1.002	0.215	4.66	<b>&lt; .001</b>
Einkommen – 2(b)								
Intercept	-2.177	0.225	-9.67	<b>&lt; .001</b>	-2.799	0.520	-5.39	<b>&lt; .001</b>
Einkommen	0.887	0.130	6.84	<b>&lt; .001</b>	1.235	0.251	4.92	<b>&lt; .001</b>
Körpergröße – 3(a)								
Intercept	-78.229	7.148	-10.94	<b>&lt; .001</b>	-77.538	13.963	-5.55	<b>&lt; .001</b>
Größe	0.464	0.042	10.93	<b>&lt; .001</b>	0.432	0.077	5.58	<b>&lt; .001</b>
Körpergröße – 3(b)								
Intercept	-91.071	8.509	-10.70	<b>&lt; .001</b>	-73.951	13.280	-5.57	<b>&lt; .001</b>
Größe	0.539	0.050	10.69	<b>&lt; .001</b>	0.413	0.074	5.60	<b>&lt; .001</b>
Körpergewicht – 4(a)								
Intercept	-11.030	1.242	-8.88	<b>&lt; .001</b>	-14.630	2.752	-5.32	<b>&lt; .001</b>
Gewicht	0.187	0.020	9.13	<b>&lt; .001</b>	0.185	0.035	5.31	<b>&lt; .001</b>
Körpergewicht – 4(b)								
Intercept	-12.146	1.093	-11.12	<b>&lt; .001</b>	-12.220	2.271	-5.38	<b>&lt; .001</b>
Gewicht	0.173	0.016	10.71	<b>&lt; .001</b>	0.137	0.027	5.14	<b>&lt; .001</b>

*Anmerkungen.* $\beta$  = Regressionsgewicht; SE = Standardfehler.<sup>1</sup> Dichotome Antwortskala.

(a) Egozentrische Perspektive; (b) Allozentrische Perspektive.

Signifikante Ergebnisse ( $p \leq .05$ ) sind fett gedruckt.

Tabelle 3.3.11.

*Ergebnisse Regressionsanalysen mit Bildungsniveau für Frauen und Männer – Bedingung 5PS<sup>1</sup>*

Merkmal und Item-Nr.	Frauen			Männer		
	$\beta$	$z$	$p$	$\beta$	$z$	$p$
Alter – 1(a)						
Reale Werte	0.057	4.73	<b>&lt; .001</b>	0.099	4.31	<b>&lt; .001</b>
Kein Abitur	0.054	0.28	.78	0.439	1.18	.24
Reale Werte : Kein Abitur	-0.013	-0.77	.44	-0.003	-0.09	.92
Alter – 1(b)						
Reale Werte	0.038	3.24	<b>.001</b>	0.055	2.84	<b>.004</b>
Kein Abitur	0.168	0.88	.38	-0.016	-0.05	.96
Reale Werte : Kein Abitur	-0.002	-0.12	.91	-0.006	-0.20	.84
Einkommen – 2(a)						
Reale Werte	0.998	7.93	<b>&lt; .001</b>	0.307	3.10	<b>.002</b>
Kein Abitur	-0.244	-1.31	.19	-0.028	-0.08	.94
Reale Werte : Kein Abitur	0.620	2.48	<b>.01</b>	0.099	0.58	.56
Einkommen – 2(b)						
Reale Werte	1.313	9.54	<b>&lt; .001</b>	1.475	6.96	<b>&lt; .001</b>
Kein Abitur	0.062	0.34	.74	-0.405	-0.99	.32
Reale Werte : Kein Abitur	0.320	1.28	.20	-0.261	-0.81	.42
Körpergröße – 3(a)						
Reale Werte	0.358	14.35	<b>&lt; .001</b>	0.354	8.01	<b>&lt; .001</b>
Kein Abitur	0.052	0.27	.79	0.107	0.28	.78
Reale Werte : Kein Abitur	-0.036	-1.17	.24	-0.094	-1.54	.12
Körpergröße – 3(b)						
Reale Werte	0.426	15.36	<b>&lt; .001</b>	0.343	8.10	<b>&lt; .001</b>
Kein Abitur	-0.007	-0.03	.97	0.142	0.38	.70
Reale Werte : Kein Abitur	-0.047	-1.39	.17	-0.054	-0.85	.40
Körpergewicht – 4(a)						
Reale Werte	0.125	11.66	<b>&lt; .001</b>	0.151	7.03	<b>&lt; .001</b>
Kein Abitur	0.078	0.40	.69	0.143	0.38	.70
Reale Werte : Kein Abitur	-0.027	-1.82	.07	-0.047	-1.78	.08
Körpergewicht – 4(b)						
Reale Werte	0.128	12.83	<b>&lt; .001</b>	0.115	6.25	<b>&lt; .001</b>
Kein Abitur	-0.184	-0.96	.34	0.161	0.45	.65
Reale Werte : Kein Abitur	0.011	0.84	.40	-0.026	-1.16	.25

*Anmerkungen.* $\beta$  = Regressionsgewicht; Reale Werte = Effekt für den Prädiktor *reale Werte* in der Referenzgruppe *Abitur*.<sup>1</sup> Fünfstufige Antwortskala.

(a) Egozentrische Perspektive; (b) Allozentrische Perspektive.

Signifikante Ergebnisse ( $p \leq .05$ ) sind fett gedruckt.

Tabelle 3.3.12.

*Ergebnisse Regressionsanalysen mit Bildungsniveau für Frauen und Männer – Bedingung 4PS<sup>1</sup>*

Merkmal und Item-Nr.	Frauen			Männer		
	$\beta$	$z$	$p$	$\beta$	$z$	$p$
Alter – 1(a)						
Reale Werte	0.044	3.79	<b>&lt; .001</b>	0.078	3.32	<b>&lt; .001</b>
Kein Abitur	0.002	0.01	.99	-0.442	-1.05	.29
Reale Werte : Kein Abitur	0.020	0.99	.32	-0.006	-0.17	.86
Alter – 1(b)						
Reale Werte	0.041	3.80	<b>&lt; .001</b>	-0.004	-0.19	.85
Kein Abitur	0.413	2.09	<b>.04</b>	-0.749	-1.76	.07
Reale Werte : Kein Abitur	-0.032	-1.61	.11	0.052	1.43	.15
Einkommen – 2(a)						
Reale Werte	1.302	10.13	<b>&lt; .001</b>	1.064	5.53	<b>&lt; .001</b>
Kein Abitur	-0.066	-0.33	.74	-0.378	-0.91	.36
Reale Werte : Kein Abitur	0.244	1.00	.32	0.429	1.08	.28
Einkommen – 2(b)						
Reale Werte	1.516	10.93	<b>&lt; .001</b>	1.223	5.70	<b>&lt; .001</b>
Kein Abitur	0.134	0.69	.49	0.426	0.98	.33
Reale Werte : Kein Abitur	0.030	0.12	.90	1.687	3.11	<b>.002</b>
Körpergröße – 3(a)						
Reale Werte	0.332	13.57	<b>&lt; .001</b>	0.313	6.71	<b>&lt; .001</b>
Kein Abitur	0.045	0.82	.82	-0.090	-0.20	.84
Reale Werte : Kein Abitur	0.038	0.30	.30	-0.132	-2.39	<b>.02</b>
Körpergröße – 3(b)						
Reale Werte	0.355	14.05	<b>&lt; .001</b>	0.307	6.71	<b>&lt; .001</b>
Kein Abitur	-0.061	-0.31	.76	0.231	0.55	.58
Reale Werte : Kein Abitur	<0.000	0.01	.99	-0.087	-1.53	.13
Körpergewicht – 4(a)						
Reale Werte	0.127	10.83	<b>&lt; .001</b>	0.127	5.69	<b>&lt; .001</b>
Kein Abitur	-0.060	-0.29	.77	0.765	1.68	.09
Reale Werte : Kein Abitur	0.012	0.66	.51	0.078	1.94	.06
Körpergewicht – 4(b)						
Reale Werte	0.138	12.08	<b>&lt; .001</b>	0.078	4.46	<b>&lt; .001</b>
Kein Abitur	-0.355	-1.77	.08	0.412	1.02	.31
Reale Werte : Kein Abitur	-0.012	-0.80	.42	0.005	0.19	.85

*Anmerkungen.* $\beta$  = Regressionsgewicht; Reale Werte = Effekt für den Prädiktor *reale Werte* in der Referenzgruppe *Abitur*.<sup>1</sup> Vierstufige Antwortskala.

(a) Egozentrische Perspektive; (b) Allozentrische Perspektive.

Signifikante Ergebnisse ( $p \leq .05$ ) sind fett gedruckt.

Tabelle 3.3.13.

*Ergebnisse Regressionsanalysen mit Bildungsniveau für Frauen und Männer – Bedingung 2PS<sup>1</sup>*

Merkmal und Item-Nr.	Frauen			Männer		
	$\beta$	$z$	$p$	$\beta$	$z$	$p$
Alter – 1(a)						
Intercept	-1.391	-9.14	<b>&lt; .001</b>	-1.331	-4.62	<b>&lt; .001</b>
Reale Werte	0.068	4.74	<b>&lt; .001</b>	0.026	1.16	.25
Kein Abitur	0.655	2.80	<b>.005</b>	-0.375	-0.74	.46
Reale Werte : Kein Abitur	-0.034	-1.75	.08	-0.004	-0.08	.93
Alter – 1(b)						
Intercept	-1.982	-10.48	<b>&lt; .001</b>	-0.816	-3.26	<b>.001</b>
Reale Werte	0.078	4.80	<b>&lt; .001</b>	-0.005	-0.24	.81
Kein Abitur	0.500	1.76	.08	-0.438	-1.00	.32
Reale Werte : Kein Abitur	-0.062	-2.77	<b>.006</b>	0.015	0.34	.74
Einkommen – 2(a)						
Intercept	-1.397	-9.28	<b>&lt; .001</b>	-0.594	-1.20	<b>.05</b>
Reale Werte	0.578	4.15	<b>&lt; .001</b>	0.809	3.74	<b>&lt; .001</b>
Kein Abitur	-0.263	-0.92	.36	-0.874	-1.42	.15
Reale Werte : Kein Abitur	0.909	2.62	<b>.009</b>	1.698	1.80	.07
Einkommen – 2(b)						
Intercept	-0.914	-6.74	<b>&lt; .001</b>	-0.297	-0.88	.38
Reale Werte	0.740	5.15	<b>&lt; .001</b>	1.246	4.00	<b>&lt; .001</b>
Kein Abitur	-0.150	-0.62	.54	0.088	0.18	.86
Reale Werte : Kein Abitur	0.616	1.92	.06	-0.051	-0.10	.92
Körpergröße – 3(a)						
Intercept	-0.438	-2.32	<b>.02</b>	0.758	2.12	<b>.03</b>
Reale Werte	0.532	8.61	<b>&lt; .001</b>	0.416	4.49	<b>&lt; .001</b>
Kein Abitur	0.337	1.17	.24	0.001	<0.00	1
Reale Werte : Kein Abitur	-0.150	-1.77	.07	0.045	0.27	.79
Körpergröße – 3(b)						
Intercept	-0.579	-2.89	<b>.004</b>	1.208	2.92	<b>.003</b>
Reale Werte	0.584	8.44	<b>&lt; .001</b>	0.459	4.52	<b>&lt; .001</b>
Kein Abitur	0.371	1.19	.23	-0.786	-1.30	.19
Reale Werte : Kein Abitur	-0.102	-1.01	.31	-0.109	-0.76	.45



Tabelle 3.3.13. (Fortsetzung)

*Ergebnisse Regressionsanalysen mit Bildungsniveau für Frauen und Männer – Bedingung 2PS<sup>1</sup>*

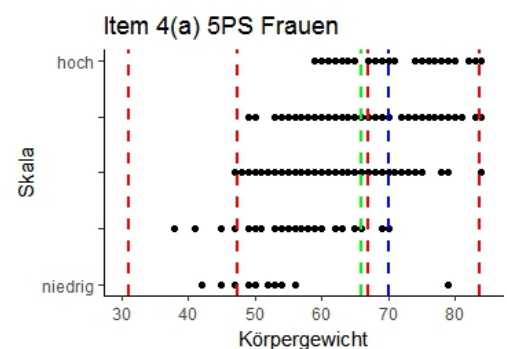
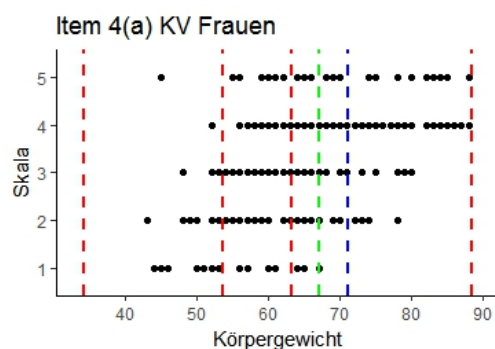
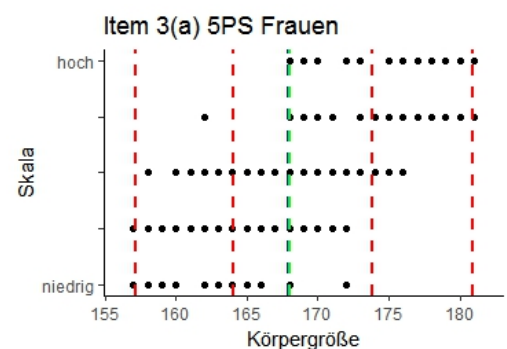
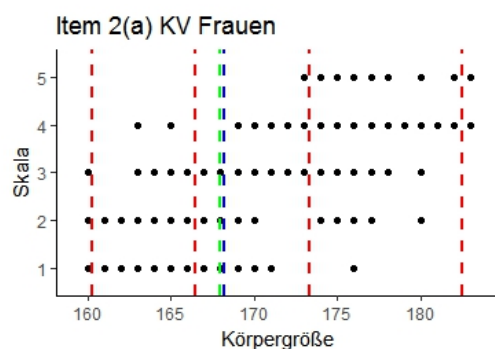
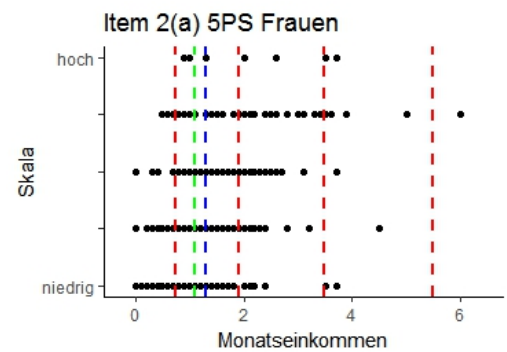
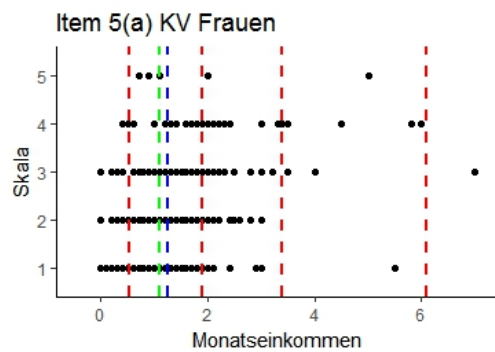
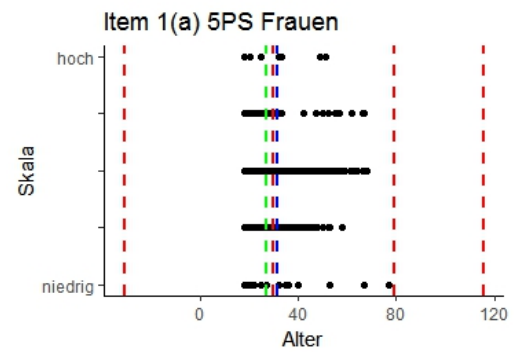
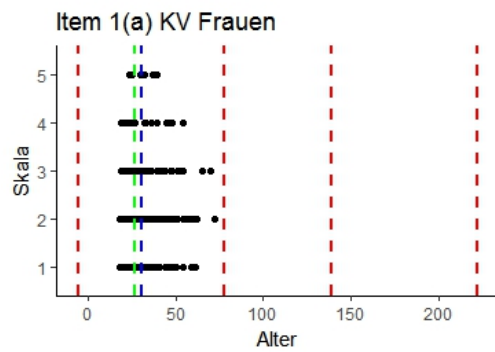
Merkmal und Item-Nr.	Frauen			Männer		
	$\beta$	$z$	$p$	$\beta$	$z$	$p$
Körpergewicht – 4(a)						
Intercept	2.027	6.89	<b>&lt; .001</b>	1.066	2.26	<b>.02</b>
Reale Werte	0.194	7.54	<b>&lt; .001</b>	0.217	4.14	<b>&lt; .001</b>
Kein Abitur	0.203	0.41	.68	-0.496	-0.79	.43
Reale Werte : Kein Abitur	-0.023	-0.53	.59	-0.065	-0.90	.37
Körpergewicht – 4(b)						
Intercept	-0.051	-0.28	.78	-0.993	-2.67	<b>.008</b>
Reale Werte	0.179	8.45	<b>&lt; .001</b>	0.170	4.23	<b>&lt; .001</b>
Kein Abitur	0.069	0.24	.81	0.422	0.80	.42
Reale Werte : Kein Abitur	-0.018	-0.56	.58	-0.073	-1.40	.16

*Anmerkungen.* $\beta$  = Regressionsgewicht; Reale Werte = Effekt für den Prädiktor *reale Werte* in der Referenzgruppe *Abitur*.<sup>1</sup> Dichotome Antwortskala.

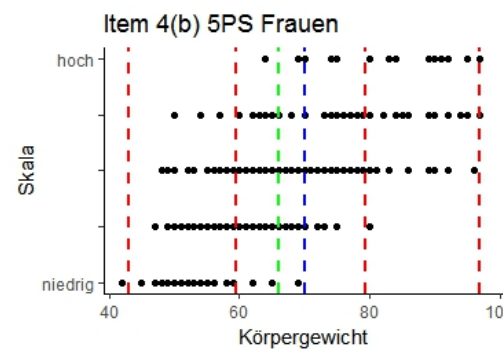
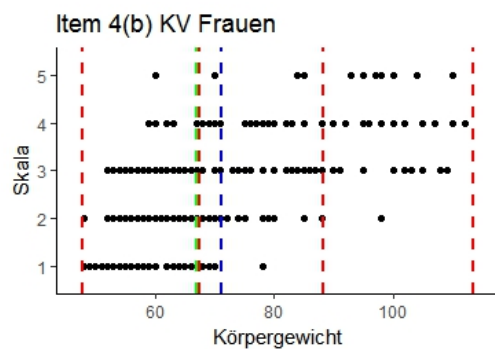
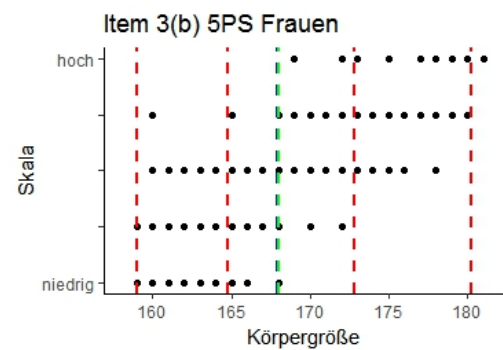
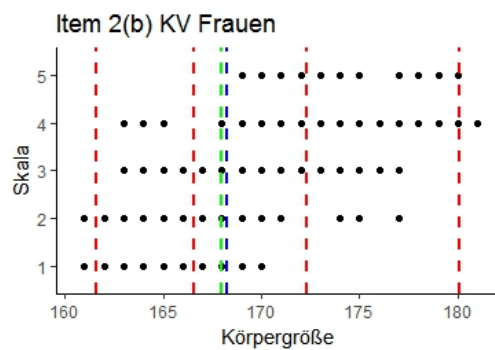
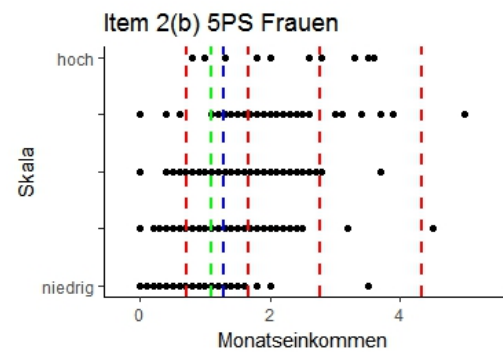
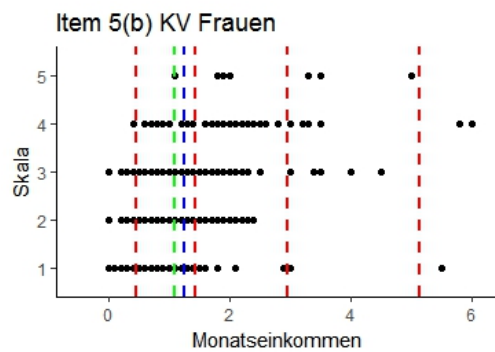
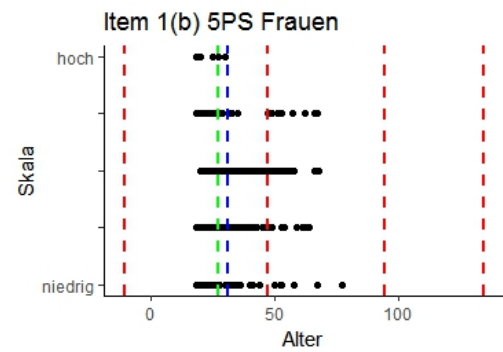
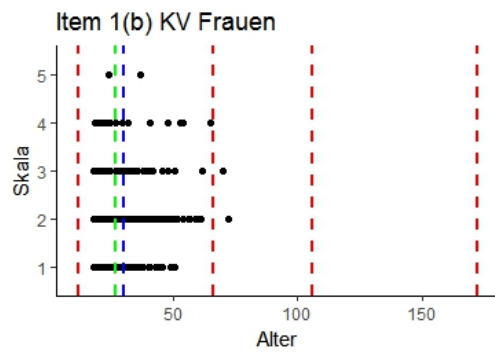
(a) Egozentrische Perspektive; (b) Allozentrische Perspektive.

Signifikante Ergebnisse ( $p \leq .05$ ) sind fett gedruckt.

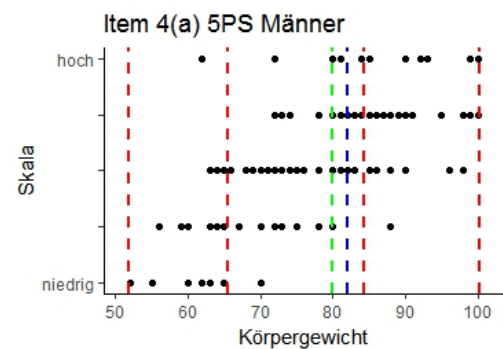
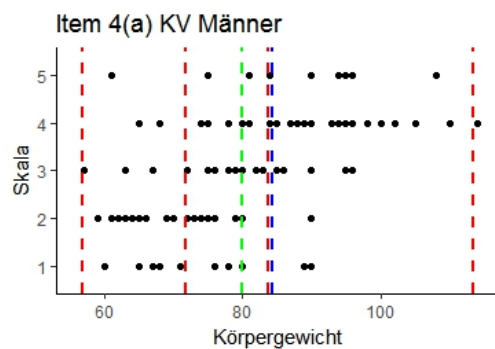
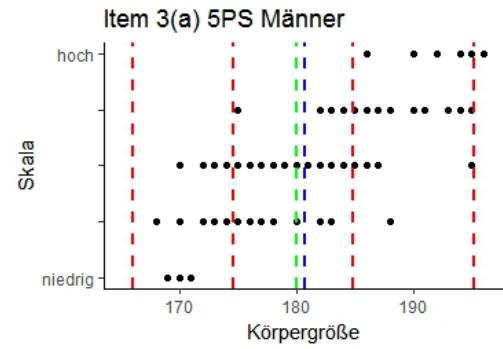
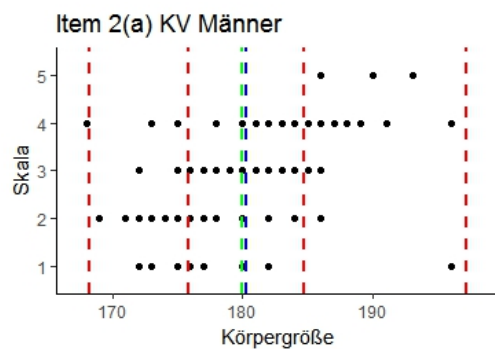
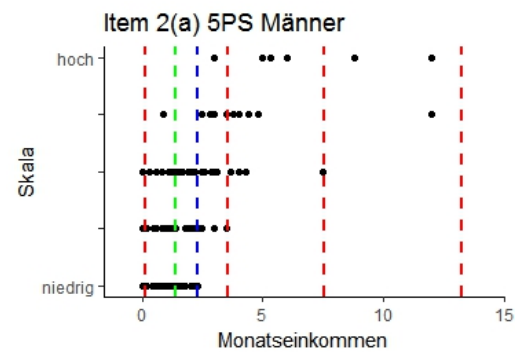
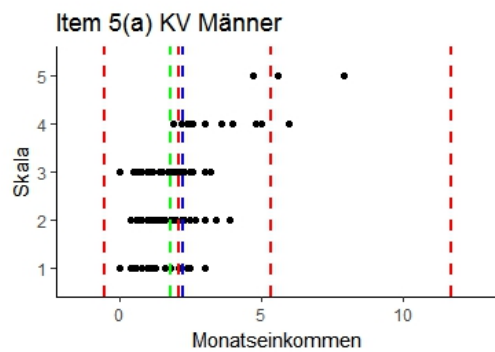
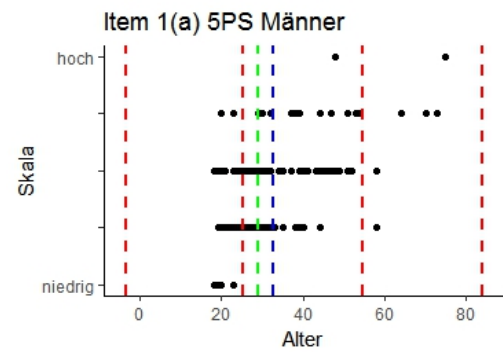
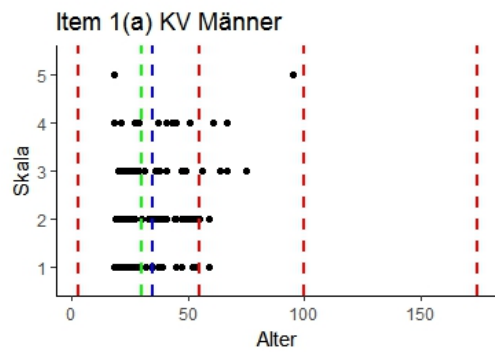
Die Abbildungen 3.3.9. bis 3.3.12. zeigen, dass sich bei beiden Geschlechtern bei den Merkmalen Alter (außer bei egozentrischer Fragenformulierung), Einkommen und Körpergröße keine auffallenden Unterschiede ergeben, wie sich die Relationen der realen Werte auf der Antwortskala in Abhängigkeit von Itemformulierung (Itemformulierung, die nicht auf einen Pol des erfragten Merkmals Bezug nimmt vs. Itemformulierung, die sich auf einen Pol des erfragten Merkmals bezieht) und Antwortskala (endpunktbenannte Antwortskala ohne numerische und verbale Verankerung vs. numerisch und verbal verankerte Likert-Skala) widerspiegeln. Beim Merkmal Körpergewicht bleiben bei beiden Geschlechtern die Relationen der realen Werte tendenziell unter Bedingung 5PS besser erhalten (nicht auf einen Pol Bezug nehmende Itemformulierung, endpunktbenannte Antwortskala ohne numerische und verbale Verankerung).



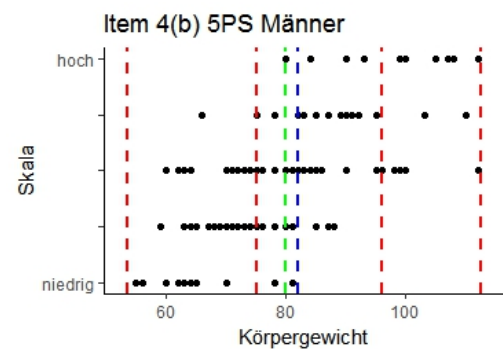
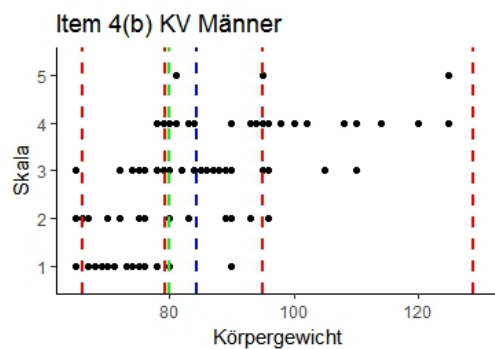
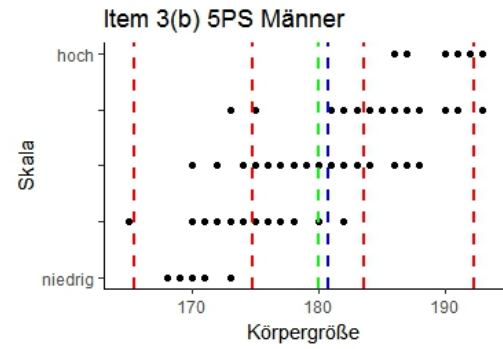
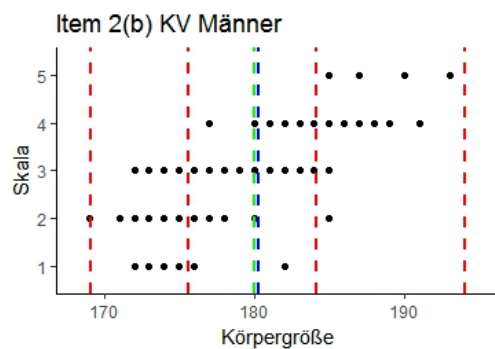
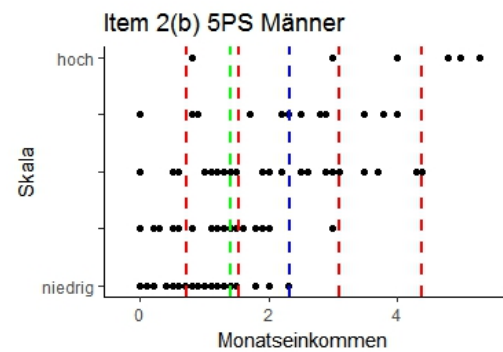
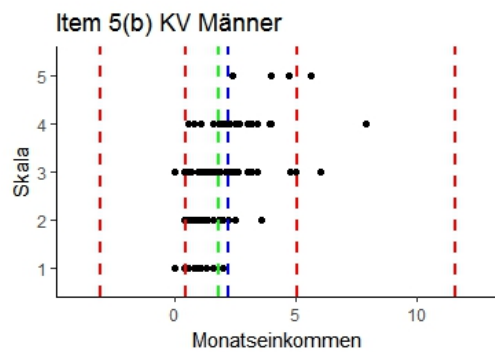
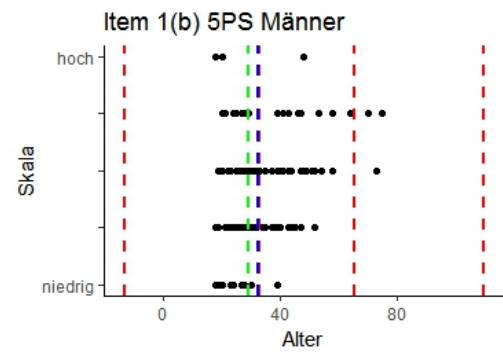
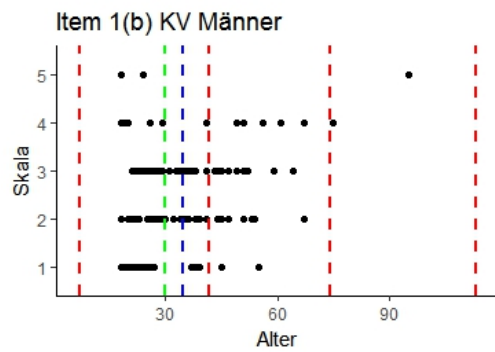
**Abbildung 3.3.9.** Regression Antwortskala und reale Werte für die Merkmale Alter, Einkommen, Körpergröße und Körpergewicht für Frauen – Egozentrische Perspektive. Alter in Jahren; Einkommen in Tausend-Euro-Werten; Körpergröße in Zentimetern; Körpergewicht in Kilogramm; schwarze Punkte = gewählte Antwortkategorien; rote Linien = geschätzte Schwellen zwischen den Kategorien der Skala; blaue Linie = Mittelwert der realen Werte; grüne Linie = Median der realen Werte; KV = Anzeige keiner Merkmalsverteilung (Daten aus Eichhorn, 2016); Itemformulierung, die sich auf einen Pol des erfragten Merkmals bezieht. 5PS = Fünfstufige Antwortskala; Itemformulierung, die nicht auf einen Pol des erfragten Merkmals Bezug nimmt. (a) Egozentrische Perspektive.



**Abbildung 3.3.10.** Regression Antwortskala und reale Werte für die Merkmale Alter, Einkommen, Körpergröße und Körpergewicht für Frauen – Allozentrische Perspektive. Alter in Jahren; Einkommen in Tausend-Euro-Werten; Körpergröße in Zentimetern; Körpergewicht in Kilogramm; schwarze Punkte = gewählte Antwortkategorien; rote Linien = geschätzte Schwellen zwischen den Kategorien der Skala; blaue Linie = Mittelwert der realen Werte; grüne Linie = Median der realen Werte; KV = Anzeige keiner Merkmalsverteilung (Daten aus Eichhorn, 2016); Itemformulierung, die sich auf einen Pol des erfragten Merkmals bezieht. 5PS = Fünfstufige Antwortskala; Itemformulierung, die nicht auf einen Pol des erfragten Merkmals Bezug nimmt. (b) Allozentrische Perspektive.



**Abbildung 3.3.11.** Regression Antwortskala und reale Werte für die Merkmale Alter, Einkommen, Körpergröße und Körpergewicht für Männer – Egozentrische Perspektive. Alter in Jahren; Einkommen in Tausend-Euro-Werten; Körpergröße in Zentimetern; Körpergewicht in Kilogramm; schwarze Punkte = gewählte Antwortkategorien; rote Linien = geschätzte Schwellen zwischen den Kategorien der Skala; blaue Linie = Mittelwert der realen Werte; grüne Linie = Median der realen Werte; KV = Anzeige keiner Merkmalsverteilung (Daten aus Eichhorn, 2016); Itemformulierung, die sich auf einen Pol des erfragten Merkmals bezieht. 5PS = Fünfstufige Antwortskala; Itemformulierung, die nicht auf einen Pol des erfragten Merkmals Bezug nimmt. (a) Egozentrische Perspektive.



**Abbildung 3.3.12.** Regression Antwortskala und reale Werte für die Merkmale Alter, Einkommen, Körpergröße und Körpergewicht für Männer – Allozentrische Perspektive. Alter in Jahren; Einkommen in Tausend-Euro-Werten; Körpergröße in Zentimetern; Körpergewicht in Kilogramm; schwarze Punkte = gewählte Antwortkategorien; rote Linien = geschätzte Schwellen zwischen den Kategorien der Skala; blaue Linie = Mittelwert der realen Werte; grüne Linie = Median der realen Werte; KV = Anzeige keiner Merkmalsverteilung (Daten aus Eichhorn, 2016); Itemformulierung, die sich auf einen Pol des erfragten Merkmals bezieht.  
5PS = Fünfstufige Antwortskala; Itemformulierung, die nicht auf einen Pol des erfragten Merkmals Bezug nimmt.  
(b) Allozentrische Perspektive.

### **3.3.3 Ergebnisse im Kontext der Forschungsfragen**

*Forschungsfrage 1.* Unter Abschnitt 3.1.4 wurde die Frage aufgeworfen, ob die Relationen der realen Werte in Abhängigkeit von Itemformulierung und Antwortskala unterschiedlich gut erhalten bleiben. Bei den Merkmalen Alter (egozentrische Formulierung) und Körpergewicht werden die Relationen der realen Werte bei beiden Geschlechtern von einer neutralen, nicht auf einen Pol Bezug nehmenden Itemformulierung in Kombination mit einer endpunktbenannten Antwortskala marginal besser widergespiegelt.

*Forschungsfrage 2.* Unter Abschnitt 3.1.4 wurde die Frage aufgeworfen, bei welcher Skalenbreite die Relationen der realen Werte am besten erhalten bleiben. Weder bei Frauen noch bei Männern ergeben sich klare Hinweise darauf, dass eine Skalenbreite den anderen Skalenbreiten überlegen ist.

*Forschungsfrage 3.* Unter Abschnitt 3.1.4 wurde die Frage aufgeworfen, ob durch unterschiedliche Itemformulierung ein Perspektivenwechsel in der Selbstbewertung erreicht werden kann, der dazu führt, dass die Relationen der realen Werte zwischen den Bedingungen im Mittel unterschiedlich gut erhalten bleiben (5PS und 4PS) bzw. dass die Lagemaße der zentralen Tendenz zwischen den Bedingungen unterschiedlich gut mit dem Wendepunkt der logistischen Funktion korrespondieren (2PS). Kleine Unterschiede zeigen sich beim Merkmal Einkommen (5PS und 2PS, Männer), auffallendere Unterschiede beim Merkmal Körpergewicht (5PS, 4PS und 2PS, beide Geschlechter, v. a. Frauen).

*Forschungsfrage 4.* Unter Abschnitt 3.1.4 wurde die Frage aufgeworfen, ob die Selbstbewertung in Abhängigkeit vom Bildungsniveau unterschiedlich ausfällt. Bei allen Merkmalen unter allen Bedingungen ändert sich weder bei Frauen noch bei Männern die Stärke des Zusammenhangs zwischen den realen Werten und der Antwortskala in Abhängigkeit vom Bildungsniveau.

## **3.4 Diskussion Studie II**

### **3.4.1 Zusammenfassung der Ergebnisse**

Zielsetzung von Studie II war die Überprüfung der Frage, ob die Selbsteinschätzung abhängig von Itemformulierung (Itemformulierung, die nicht auf einen Pol des erfragten Merkmals Bezug nimmt vs. Itemformulierung, die sich auf einen Pol des erfragten Merkmals bezieht), Art der Skalenverankerung (endpunktbenannte Antwortskala ohne numerische und verbale Verankerung vs. numerisch und verbal verankerte Likert-Skala) und

Skalenbreite (5PS, 4PS und 2PS) divergiert. Ergänzend wurde überprüft, ob sich die Selbstbewertungsperspektive erneut als eine den Referenzrahmen der Selbstbewertung bestimmende Variable zeigt und ob die Selbstbewertung abhängig vom Bildungsniveau unterschiedlich ausfällt. Untersucht wurde, unter welchen dieser Bedingungen die Relationen der realen Werte im Mittel am besten erhalten bleiben.

Die Relationen der realen Werte blieben bei beiden Geschlechtern bei Körpergewicht und Alter (egozentrische Formulierung) bei neutraler, nicht auf einen Pol bezogener Itemformulierung zusammen mit einer endpunktbenannten Antwortskala im Mittel marginal besser erhalten. Bei allen anderen Merkmalen zeigten sich diesbezüglich keine bedeutenden Unterschiede. Es gab keine Hinweise darauf, dass eine Skalenbreite den anderen Skalenbreiten in Bezug auf die vorliegende Fragestellung überlegen ist.

Die Selbstbewertungsperspektive konnte als eine weitere den Referenzrahmen der Selbstbewertung bestimmende Variable vor allem beim Merkmal Körpergewicht (Bedingung 5PS, 4PS und 2PS, beide Geschlechter, v. a. bei Frauen), aber auch beim Merkmal Einkommen (Bedingung 5PS und 2PS, Männer) beobachtet werden. Die Selbstbewertung blieb in allen Merkmalen unabhängig vom Bildungsniveau.

Die Interpretation der Ergebnisse bezieht sich vornehmlich auf die Hauptbefunde.

### **3.4.2 Interpretation und Implikationen der Ergebnisse**

Wie unter Abschnitt 3.1.1 illustriert, kann die Formulierung einer Frage maßgeblich das Antwortverhalten bestimmen (z. B. Schuldt, Roh & Schwarz, 2015; Schuman & Presser, 1981; Schwarz, 1999). Deshalb setzte sich Payne (1951) bereits vor über sechzig Jahren in seinem Standardwerk *The Art of Asking Questions* eingehend mit der Konzeption verständlich und präzise formulierter Items auseinander. Autoren von Lehrbüchern empirisch-sozialwissenschaftlicher Forschung beziehen sich häufig auf seine Empfehlungen. Dabei geben sie, unter Betonung der Relevanz für Reliabilität und Validität, umfangreiche Richtlinien zur Itemformulierung vor (z. B. Aiken & Groth-Marnat, 2006; Bryman, 2016; Bühner, 2011; Döring & Bortz, 2016; Eid & Schmidt, 2014; Faulbaum, Prüfer & Rexroth, 2009; Jonkisz et al., 2012; Miller & Lovler, 2016; Murphy & Davidshofer, 2001; Porst, 2014; Pospeschill, 2010; Rost, 2004). Aspekte bezüglich negativ gepolter und/oder negativ formulierter Items wurden unter Abschnitt 3.1.2 dargestellt. In Studie II zeigten sich keine deutlichen Unterschiede zwischen einer Itemformulierung, die nicht auf einen Pol des

erfragten Merkmals Bezug nimmt und einer Itemformulierung, die sich auf einen Pol des erfragten Merkmals bezieht. Die Relationen der realen Werte blieben in beiden Fällen weitestgehend gleich erhalten. Damit reihen sich die Ergebnisse von Studie II in bestehende Vorbefunde ein, bei denen die Auflösung negativer Itempolungen ebenfalls nur zu marginal veränderten psychometrischen Kennwerten führte. Beispielsweise verglich eine Studie das aktuelle englische *NEO-Persönlichkeitsinventar (NEO-PI-3)* mit seinem Vorgänger, dem *NEO-Persönlichkeitsinventar, revidierte Fassung (NEO-PI-R)* (McCrae, Costa & Martin, 2005). In der aktuellen Version wurden 37 Items des *NEO-PI-R* vereinfacht, da diese wiederholt als schwer verständlich eingestuft wurden. Die Neuformulierung führte nur zu marginalen Verbesserungen in den Reliabilitätsschätzungen. Pargent, Hilbert, Eichhorn und Bühner (2018) untersuchten systematisch den Einfluss der Richtlinien zur Itemformulierung auf die psychometrischen Eigenschaften eines Fragebogens anhand des *NEO-Fünf-Faktoren-Inventars (NEO-FFI)* von Borkenau und Ostendorf (2008). Dafür wurden die Items des *NEO-FFI* einmal entlang und einmal entgegen den Richtlinien zur Itemformulierung umformuliert. Die entsprechenden Gut- und Schlecht-Versionen wurden jeweils einer großen Stichprobe vorgelegt und die ermittelten psychometrischen Kennwerte mit denen der Originalversion des *NEO-FFI* verglichen. Dabei wiesen die beiden Alternativversionen vergleichbare psychometrische Eigenschaften auf wie die Originalversion. Eine Folgestudie, die die Items der deutschen Version des *Positive and Negative Affect Schedule* (Krohne, Egloff, Kohlmann & Tausch, 1996) ebenfalls entlang und entgegen der Richtlinien zur Itemformulierung umformulierte, zeigte ähnliche Ergebnisse (Goretzko, Pargent, Sust & Bühner, 2018). Auch dort blieben die psychometrischen Kennwerte der Gut- und Schlecht-Versionen im Vergleich zu den Originalitems weitestgehend unverändert. In Studie II wurden die Items nicht in Schlecht-Versionen umformuliert. In der veränderten Formulierung wurde lediglich der Bezugspunkt der Items verändert. Keine der Formulierungsvarianten war missverständlich, wobei dies strenggenommen mithilfe eines kognitiven Interviews überprüft werden müsste. Nachdem sogar schwer verständliche Items kaum Einfluss auf psychometrische Eigenschaften haben, ist naheliegend, dass sich die vorgenommenen Veränderungen der Itemformulierung ebenfalls nicht in den untersuchten Parametern niederschlugen.

Bei den Merkmalen Alter (egozentrische Formulierung) und Körpergewicht (egozentrische und allozentrische Formulierung) bildete die endpunktbenannte



Antwortskala die Relationen der realen Werte marginal besser ab. Sie kam im Vergleich zur fünfstufigen Likert-Skala dem Intervallskalenniveau näher. Dies könnte im Einklang mit der Annahme stehen, dass endpunktbenannte Skalen von den Befragten eher als äquidistant wahrgenommen werden und mehr der Idee einer intervallskalierten Skala entsprechen (Mummendey & Grau, 2014; Porst, 2014; Rohrmann, 1978). Warum sich dies nur bei den Merkmalen Alter und Körpergewicht zeigte, bleibt damit allerdings unbeantwortet. Eine plausiblere Erklärung könnte aufgrund der geringen Unterschiede sein, dass diese eher durch Stichprobenfehler als durch systematische Einflüsse zustande kamen.

Bei allen Merkmalen unter allen Bedingungen ergaben sich weder bei Frauen noch bei Männern klare Hinweise darauf, dass eine der Skalenbreiten die Relationen der realen Werte besser erhält als die anderen untersuchten Skalenbreiten. Beispielsweise unterschied sich die Stärke des Zusammenhangs zwischen gewählter Antwortkategorie und realen Werten nicht systematisch zwischen den verschiedenen Skalenbreiten (vgl. Tabelle 3.3.7.). Diese Befunde reißen sich in die inkonsistente Studienlage zur Skalenbreite ein. Wie unter Abschnitt 3.1.3 beschrieben, können Studien einen Zusammenhang zwischen Skalenbreite und psychometrischen Eigenschaften sowohl zeigen (z. B. Lozano et al., 2008; Maydeu-Olivares et al., 2009; Ooster, 1989; Preston & Colman, 2000; Weng, 2004) als auch widerlegen (z. B. Drake et al., 2014; Matell & Jacoby, 1971). Selbst die postulierte Einschränkung von Reliabilität und Validität bei dichotomen Items im Vergleich zu polytomen Items (z. B. Preston & Colman, 2000; Weng, 2004) ist fraglich. So verglichen Finn, Ben-Porath und Tellegen (2015) das dichotome Antwortformat des *Minnesota Multiphasic Personality Inventory-2-Restructured Form* (Ben-Porath & Tellegen, 2008/2011) mit einem neuen, vierstufigen Antwortformat. Dabei zeigten sich zwischen beiden Versionen nur marginale Reliabilitätsunterschiede und eine gleichbleibende externe Kriteriumsvalidität. Möglicherweise wird der Einfluss der Skalenbreite auf die psychometrischen Eigenschaften überschätzt oder schlägt sich in den angesetzten Untersuchungsparametern von Studie II nicht deutlich genug nieder.

Unter Umständen spiegeln die Ergebnisse von Studie II auch die grundsätzliche Ambiguität der Forschung in Bezug auf Antwortskalen und Itemformulierung wider. So geht nahezu jeder Vorteil eines bestimmten Skalentyps mit Nachteilen einher. Beispielsweise reduzieren verbalisierte Skalen zwar extreme Antwortstile und fehlerhaftes

Antwortverhalten, begünstigen jedoch Akquieszenz (Weijters et al., 2010). Die Studienergebnisse zur Itemformulierung sind gleichermaßen uneindeutig. So kann einerseits die Problematik negativ formulierter Items in Bezug auf psychometrische Eigenschaften aufgezeigt werden (Barnette, 2000; Hinz et al., 2003; Schriesheim et al., 1991; Schriesheim & Hill, 1981; Wang et al., 2015). Andererseits kann es sein, dass negativ formulierte Items konsistenter beantwortet werden als positiv formulierte Items (Krebs & Matschinger, 1993). Eine Erklärung für diese Widersprüchlichkeit könnte sein, dass Messungen mittels psychologischer Fragebögen hohe Messfehler enthalten, sodass Unterschiede in der Itemformulierung und/oder im Antwortformat vom allgemeinen Rauschen überlagert werden.

Durch unterschiedliche Itemformulierung konnte bei den Merkmalen Einkommen (5PS und 2PS, Männer) und Körpergewicht (5PS, 4PS und 2PS, beide Geschlechter, v. a. bei Frauen) ein Perspektivenwechsel in der Selbstbewertung erreicht werden, jedoch nicht bei den Merkmalen Alter und Körpergröße. Damit wurden die Ergebnisse in Eichhorn (2016) für die Merkmale Einkommen, Körpergewicht und Körpergröße bestätigt. Als möglichen Erklärungsansatz diskutierte Eichhorn (2016), dass die allozentrische Itemformulierung die Subjektivität bei der Merkmalswahrnehmung in den Hintergrund drängt und dadurch eine objektivere und realistischere Selbsteinschätzung ermöglicht. Im Gegensatz dazu könnte die egozentrische Itemformulierung dazu führen, dass die Subjektivität bei der Einschätzung stärker in den Vordergrund gerät, weshalb die Relationen der realen Werte im Mittel schlechter erhalten bleiben. Besonders relevant könnte dies bei Merkmalen sein, deren Selbstbewertung stark durch subjektive Faktoren beeinflusst wird, wie beispielsweise Körpergewicht bei Frauen.

Studien zeigen, dass Frauen in westlichen Kulturen im Mittel ein ungünstigeres Körperbild haben als Männer, weshalb sie u. a. ihr Gewicht höher einschätzen (Jackson, Sullivan & Rostker, 1988). Bei egozentrischer Formulierung zeigte sich dieses geschlechtsspezifische Antwortmuster stärker als bei allozentrischer Formulierung. Es könnte also sein, dass die allozentrische Perspektive dazu führte, dass das subjektive, ungünstig verzerrte Körperbild in den Hintergrund trat und eher das objektive Körpergewicht bewertet wurde.

Beim Merkmal Körpergröße ergaben sich zwischen beiden Formulierungsperspektiven kaum Unterschiede. Folglich müsste Körpergröße ein Merkmal sein, dessen Selbsteinschätzung kaum subjektiven Einflussfaktoren unterliegt. Diese Annahme wird dadurch untermauert, dass zur Bestimmung des wahrgenommenen Körperbildes vor allem Gewicht, Figur und allgemeine Attraktivität herangezogen werden und seltener Körpergröße (Montepare, 1996a; Jackson et al., 1988). Obwohl kleine Frauen von Männern attraktiver bewertet werden als größere Frauen (Shepperd & Strathman, 1989), ist der entscheidende Faktor bei der Beurteilung der Körperattraktivität nicht die Körpergröße, sondern ein für die jeweilige Körpergröße eher niedriges Körpergewicht (Swami & Furnham, 2008). So werden Frauen mit einem BMI im unteren Normbereich von 19 bis 20 am attraktivsten beurteilt (Swami & Furnham, 2008). Frauen geben zwar bei der Partnersuche eine allgemeine Präferenz für große Männer an, mit denen sie sich auch eher verabreden würden, bewerten aber große Männer nicht attraktiver (Shepperd & Strathman, 1989). Bei beiden Geschlechtern scheint allein Körpergröße für die Wahrnehmung des Körperbildes keine saliente Bezugsgröße zu sein. Möglicherweise ist Körpergröße ein Merkmal, das weniger psychologische Bedeutung transportiert als andere Merkmale. Dies könnte dazu führen, dass subjektiv empfundene und objektive Körpergröße nur gering divergieren, weshalb sich der Perspektivenwechsel in der Itemformulierung in den untersuchten Parametern kaum niederschlägt.

Konträr zu den Ergebnissen in Eichhorn (2016) konnte beim Merkmal Alter über die Itemformulierung kein innerer Perspektivenwechsel in der Selbstbewertung induziert werden. Jedoch war bereits in Eichhorn (2016) das Merkmal Alter durch unplausible Parameterschätzungen gekennzeichnet. Beispielsweise wurden Schwellen mit negativem Vorzeichen geschätzt oder bei einem Alter zwischen 200 und 300 Jahren (vgl. Abbildung 3.3.9.). Da in allen Gruppen alle Antwortkategorien besetzt sind, kamen diese Schwellenparameter vermutlich aufgrund geringer Zusammenhänge zwischen den realen Werten und der Antwortskala zustande. Die in Eichhorn (2016) vorgenommene Ergebnisinterpretation in Bezug auf den inneren Perspektivenwechsel ist deshalb mit Vorsicht zu betrachten. In Studie II fallen die Parameterschätzungen gleichermaßen unplausibel aus. Beispielsweise wurden Schwellen bei einem Alter zwischen 400 und 500 Jahren geschätzt oder ebenfalls mit negativem Vorzeichen (vgl. Abbildung 3.3.2.), obwohl in

allen Gruppen alle Antwortkategorien besetzt sind mit Ausnahme von Item 1(a), Bedingung 4PS bei Männern. Des Weiteren zeigten sich teilweise Nullzusammenhänge zwischen den realen Werten und der Antwortskala (vgl. Tabelle 3.3.7.). Nur bei Frauen war ein konstanter, wenn auch relativ schwacher Zusammenhang zwischen den realen Werten und der Antwortskala zu beobachten. Eine Erklärung dafür könnte sein, dass die Selbstbewertung vornehmlich auf Basis des subjektiv empfundenen Alters vorgenommen wurde. Dieses entspricht häufig nicht dem durch die realen Werte abgebildeten chronologischen Alter (Montepare & Lachman, 1989; Schwall, 2012). So nehmen sich jüngere Personen im Mittel gezielt älter wahr als sie sind und ältere Personen im Mittel gezielt jünger (Barrett, 2005; Montepare, 1996b; Montepare & Lachman, 1989). Überdies kann das subjektiv empfundene Alter eine weitaus größere Prädiktionsgüte für das Verhalten einer Person haben als das chronologische Alter. So kauft eine 65-jährige Person mit einem empfundenen Alter von 45 Jahren eher Produkte, die diesem Alter entsprechen als zu ihrem chronologischen Alter passende (Schwall, 2012). Subjektiv empfundenes Alter ist wiederum ein mehrdimensionales Konstrukt, dem biologische, soziale und psychologische Prozesse zugrunde liegen. Distinkte Dimensionen spiegeln dabei das wahrgenommene Alter von Personen in Bezug auf ihr Aussehen, Verhalten, Handeln, Gefühl sowie ihre Interessen wider (Kastenbaum, Derbin, Sabatini & Artt, 1972). Die Selbsteinschätzung des Alters kann deshalb zwischen Personen gleichen chronologischen Alters stark variieren, da dabei jeder einen individuellen, subjektiven und kontextabhängigen Bezugsrahmen herstellt (Schwall, 2012). Dadurch würden die Antwortkategorien einer Ratingskala nicht immer dieselben realen Werte abbilden, was sich in schwachen systematischen Zusammenhängen niederschlagen könnte.

Gegen diese Argumentation spricht, dass das empfundene Alter zwar vom chronologischen Alter mehr oder weniger stark abweichen kann, beide Konstrukte dennoch eng zusammenhängen ( $r = .81$ ; Cleveland & Shore, 1992, S. 474). Auch wenn die Selbsteinschätzung auf Basis des empfundenen Alters erfolgt, hat das chronologische Alter die höchste Prädiktionsgüte für das empfundene Alter (Schwall, 2012). Die beobachteten schwachen Zusammenhänge bzw. Nullzusammenhänge zwischen chronologischem Alter und Antwortskala erscheinen deshalb unplausibel. Eventuell ist Alter ein Merkmal, das verglichen mit den anderen untersuchten Merkmalen die höchste Komplexität und

Mehrdimensionalität aufweist. Dies könnte zu einer besonders großen Messfehleranfälligkeit führen. So fielen sowohl in Eichhorn (2016) als auch in Studie II die Parameterschätzungen vor allem für das Merkmal Alter in höchstem Maße unplausibel aus. Möglicherweise ist bei dessen Erfassung der Fehlervarianzanteil so hoch, dass sich Unterschiede in der Itemformulierung je nach Stichprobe einmal zeigen und ein andermal im allgemeinen Rauschen untergehen. Womöglich sind deshalb die Ergebnisse beim Merkmal Alter generell mit Vorsicht zu interpretieren.

Bei Frauen zeigten die Brant-Tests überwiegend eine Modellverletzung, bei Männern eine Modellgültigkeit. Die Teststärke des Brant-Tests steigt mit zunehmender Stichprobengröße (Harrell, 2015). Deshalb sind die Ergebnisse bei Männern mit Vorsicht zu interpretieren, da möglicherweise die vorliegenden Stichprobengrößen zur Erkennung einer Modellverletzung nicht ausreichend sind (Bedingung 5PS:  $n = 131$ ; Bedingung 4PS:  $n = 113$ ). Eine Modellgültigkeit bei Männern kann daher nicht gesichert angenommen werden. Auch wenn das Proportional Odds Modell für die vorliegende Fragestellung grundsätzlich das theoretisch sinnvolle Modell ist (Agresti, 2010; Agresti, 2013; Harrell, 2015; Tutz, 2012), kann die Modellgültigkeit bei beiden Geschlechtern und für alle Merkmale vermutlich nicht angenommen werden.

Die Stärke des Zusammenhangs zwischen den realen Werten und der Antwortskala änderte sich nicht in Abhängigkeit vom Bildungsniveau. Damit konnten die Ergebnisse aus Studie I bestätigt werden. Mögliche Erklärungsansätze wurden bereits unter Abschnitt 2.4.2 diskutiert.

### **3.4.3 Limitationen und Ausblick**

Die Teilnehmerzahl von insgesamt 1 769 Personen ist zufriedenstellend. Die drei Substichproben sind bezüglich ihrer demographischen Verteilung weitestgehend vergleichbar. Effekte aufgrund zu großer Unterschiedlichkeit können daher vernachlässigt werden. Jedoch zeigte sich ein unausgewogenes Verhältnis bei den Merkmalen Geschlecht (Frauenanteil: 79.4 %), Bildung (Allgemeine Hochschulreife: 66.1 %) und Alter (linkssteile Altersverteilung;  $M = 32$ ). Da Studien darauf hinweisen, dass das Antwortverhalten u. a. in Abhängigkeit von Geschlecht (Hall & Matsumoto, 2004), Bildung (Meisenberg & Williams, 2008) und Alter (Weijters, Geuens & Schillewaert, 2010) variieren kann, ist die Generalisierbarkeit der Ergebnisse eingeschränkt. Eine Replikation der Studie an einer in

Bezug auf Geschlecht, Alter und Bildungsniveau repräsentativen Stichprobe wäre erstrebenswert.

In Studie II wurde die Itemformulierung stark verändert, weshalb das Antwortformat mit angepasst werden musste. Die Beibehaltung einer fünfstufigen Likert-Skala wäre nicht praktikabel gewesen. Deshalb ist nicht auszuschließen, dass sich potenzielle Effekte auf beiden Seiten gegenseitig beeinflussten oder überlagerten. Möglicherweise wurden die Vorteile einer veränderten Itemformulierung durch die Nachteile einer veränderten Antwortskala aufgehoben oder umgekehrt. Bei weiterführenden Studien sollte deswegen darauf geachtet werden, dass Itemformulierung und Antwortformat Schritt für Schritt variiert werden. So könnte bei identischer Itemformulierung zunächst das Antwortformat verändert werden. Dadurch könnten Effekte, die auf das Skalenformat zurückzuführen sind, besser von Effekten, die auf die Itemformulierung zurückgehen, getrennt werden. In einem nächsten Schritt könnten sodann Veränderungen an der Itemformulierung bei identischem Antwortformat vorgenommen werden, um wiederum Effekte der Itemformulierung von Effekten des Antwortformats zu separieren.

#### **3.4.4 Fazit**

In Studie II hatten Itempolung, Art der Skalenverankerung und Skalenbreite keinen oder nur marginalen Einfluss darauf, wie gut die Relationen der realen Werte auf der Antwortskala erhalten bleiben. Die beobachteten Unterschiede sind nicht eindeutig genug, um sie systematischen Einflüssen zuschreiben zu können.

Durch unterschiedliche Itemformulierung konnte bei den Merkmalen Einkommen (Männer) und Körpergewicht (beide Geschlechter) ein Perspektivenwechsel in der Selbstbewertung erreicht werden, jedoch nicht bei den Merkmalen Alter und Körpergröße. Die Selbstbewertungsperspektive kam eher bei Merkmalen zum Tragen, die stark von subjektiven Faktoren beeinflusst werden. Sollen die erhobenen Werte möglichst nah an den realen Werten liegen, scheint die allozentrische Formulierungsvariante der egozentrischen überlegen zu sein.

Für das Merkmal Alter fielen die Parameterschätzungen wie bereits in Eichhorn (2016) äußerst unplausibel aus, weshalb die Ergebnisse mit Vorsicht interpretiert werden sollten. Möglicherweise ist Alter ein Merkmal mit so hoher Komplexität und Mehrdimensionalität,

dass dessen Erfassung mit erheblichen Messfehlern behaftet ist und Änderungen in der Itemformulierung vom allgemeinen Rauschen überlagert werden.

Das Bildungsniveau hatte erneut bei keinem Merkmal Einfluss auf die Stärke des Zusammenhangs zwischen den realen Werten und der Antwortskala. Damit bestätigten sich die Ergebnisse aus Studie I.

## **4. Studie III: Konstruktion einer psychometrischen Skala für physikalische Merkmale**

### **4.1 Theorie Studie III**

#### **4.1.1 Einführung**

Die Studien I und II beschäftigten sich mit der psychometrischen Güte einzelner Items. Zentrale Fragestellung dabei war, wie gut ein Item das zu erfassende Merkmal abbilden kann. Die Itemkonstruktion folgte dabei keiner speziellen Prozedur. Jedes Item (und seine Formulierungsvarianten) wurde direkt aus dem Messgegenstand abgeleitet, wie dies bei Single-Item-Skalen üblicherweise der Fall ist (z. B. „Ich bin groß.“ oder „Ich halte mich für groß.“ für das Merkmal Körpergröße). Single-Item-Skalen bilden keine mit dem zu erfassenden Konstrukt assoziierten Einstellungen, Innenansichten oder Verhaltensausschnitte ab, sondern entsprechen mehr oder weniger der Kerndefinition des Messgegenstandes. Beispiele dafür sind das Item „To what extent do you agree with this statement: I am a narcissist.“ (Konrath, Meier & Bushman, 2014, S. 3) zur Erfassung von Narzissmus, das Item „Have you felt depressed or sad much of the time in the past year?“ (Williams et al., 1999, S. 37) zur Erfassung von Depression oder das Item „Wie zufrieden sind Sie gegenwärtig, alles in allem, mit Ihrem Leben?“ (Beierlein, Kovaleva, László, Kemper & Rammstedt, 2014, S. 6) zur Erfassung allgemeiner Lebenszufriedenheit.

Single-Item-Skalen sind ökonomische Messinstrumente zur Erfassung psychischer Merkmale in Anwendungsfällen, die starken zeitlichen und finanziellen Restriktionen unterliegen. Dies ist beispielsweise bei soziologischen Panelbefragungen (z. B. Beierlein et al., 2014) oder klinischen Screening-Programmen der Fall (z. B. Corson, Gerrity & Dobscha, 2004; Rhew et al., 2010; Williams et al., 1999; für eine Metaanalyse vgl. Mitchell & Coyne, 2007). Aufgrund des geringen Zeitaufwandes für Bearbeitung und Auswertung lassen sie sich leicht in Umfragepanels bzw. Routineuntersuchungen integrieren. Dabei zeigt sich, dass einzelne Items bezüglich ihrer praktischen Relevanz, beispielsweise bei der Beantwortung der Frage, ob eine Depression vorliegt, ähnlich leistungsfähig sein können, wie umfangreichere Fragebögen (Michalos & Kahlke, 2010; Williams et al., 1999). Zur präzisen Messung breit angelegter psychologischer Konstrukte sind sie jedoch ungeeignet, weshalb sie in der empirischen Sozialforschung unüblich sind (Konrath et al., 2014; Michalos & Kahlke, 2010). Im Forschungskontext werden sie nur dann empfohlen, wenn es schwierig oder unmöglich ist, längere Skalen in Studien zu integrieren (Beierlein et al., 2014; Konrath



et al., 2014). Zum Einsatz kommen zumeist Fragebögen bzw. Skalen, die aus mehreren Items bestehen. Sie sind eines der wichtigsten und am häufigsten angewendeten Instrumente in der sozialwissenschaftlichen Forschung (Kallus, 2016; Mummendey & Grau, 2014).

Aufgrund der hohen Relevanz von Fragebögen wird in Studie III der Zusammenhang zwischen physikalisch messbaren Personenmerkmalen und deren Abbildung auf psychologischen Skalen untersucht. Dafür werden, den in der Fachliteratur gegebenen Empfehlungen zur Fragebogenkonstruktion folgend, drei Skalen für die physikalischen Merkmale Alter, Körpergröße und Körpergewicht entwickelt. Die Beantwortung der Items erfolgt entweder auf einer polytomen (sechsstufiges Antwortformat) oder dichotomen Antwortskala. Im Mittelpunkt der Analysen steht die Frage, wie gut die Skalen in der Lage sind, die physikalischen Merkmale abzubilden. Genauer gesagt geht es um die Frage der Validität nach Kelley (1927, S. 14), „whether a test really measures what it purports to measure“.

#### **4.1.2 Fragebogen als Messverfahren**

Die empirische Sozialforschung versteht unter einem Fragebogen ein Messverfahren, das auf testtheoretischen Konzepten basiert und psychometrischen Gütekriterien entsprechen soll (Kallus, 2016; Moosbrugger & Kelava, 2012). Dabei werden Fragen, Begriffe oder Aussagen (die sogenannten Items) konzipiert, die von den befragten Personen mithilfe festgelegter Antwortmöglichkeiten bearbeitet werden sollen (Mummendey & Grau, 2014). Davon abzugrenzen sind schriftliche Befragungen, in denen aus verschiedenen inhaltlichen Bereichen Informationen erfasst werden, wie beispielsweise biografische, medizinische oder demoskopische Daten (Kallus, 2016; Moosbrugger & Kelava, 2012). In Anlehnung an das englische Wort *scale* wird ein psychologischer Fragebogen auch häufig als *Skala* bezeichnet (Moosbrugger & Kelava, 2012). Ein testtheoretisch konstruierter Fragebogen ist nicht nur eine empiristische Aneinanderreihung von Items, sondern eine theoretisch begründete und den Prinzipien einer Testtheorie folgende Auswahl an Items. In ihrer Gesamtheit sollen diese Items ein einziges psychologisches Konstrukt, d. h. ein ihnen zugrundeliegendes theoretisches Konzept erfassen (Mummendey & Grau, 2014; Porst, 2014). In diesem Sinne wird ein Fragebogen als *psychologischer Test* klassifiziert (Lienert & Raatz, 1998). Demnach ist ein Test ein „wissenschaftliches Routineverfahren zur Erfassung eines oder mehrerer empirisch abgrenzbarer Persönlichkeitsmerkmale mit dem Ziel einer möglichst quantitativen

Aussage über den relativen Grad der individuellen Merkmalsausprägung“ (Lienert & Raatz, 1998, S. 1).

Die Entwicklung eines solchen Tests gestaltet sich dementsprechend kompliziert und fordernd (Porst, 2014). Entsprechende Literatur versucht Hilfestellung bei der Item- und Fragebogenkonstruktion zu geben. Dabei wird in älteren Publikationen die Fähigkeit zur Fragebogenkonstruktion vor allem von persönlichen Fähigkeiten und eigener Einfallskraft (Noelle, 1963) sowie von Geschick und Erfahrung (Scheuch, 1962) abhängig gemacht. Die Itemformulierung wird als Gefühlssache (Holm, 1974) oder gar als Kunstlehre (Payne, 1951) betrachtet. Im Gegensatz dazu wird in aktuellerer Literatur die Notwendigkeit allgemeiner Konstruktionsregeln und die Wichtigkeit der Umsetzung wissenschaftlicher Erkenntnisse, insbesondere aus der Kognitionspsychologie, betont (z. B. Bradburn, Sudman & Wansink, 2004; Faulbaum et al., 2009; Jacob, Heinz & Décieux, 2013; Schnell, Hill & Esser, 2013; Sudman, Bradburn & Schwarz, 1996; Sudman & Wansink, 2002). Die Richtlinien zur Itemformulierung sind ein wichtiger Teil dieser Konstruktionsregeln, die sich in Lehrbüchern zu Test- und Fragebogenkonstruktion finden. Im Rahmen dieser Richtlinien werden Empfehlungen zur Itemformulierung gegeben. Beispielsweise sollen Items eindeutig, widerspruchsfrei, klar, kurz, nicht überfordernd und grammatikalisch korrekt formuliert werden (z. B. Aiken & Groth-Marnat, 2006; Bühner, 2011; Döring & Bortz, 2016; Eid & Schmidt, 2014; Faulbaum et al., 2009; Jonkisz et al., 2012; Miller & Lovler, 2016; Murphy & Davidshofer, 2001; Porst, 2014; Pospeschill, 2010; Rost, 2004). Die Erforschung kognitiver und kommunikativer Prozesse, die der Befragungssituation zugrunde liegen, ließ die Fragebogenkonstruktion eine neue Qualität erfahren (Porst, 2014). So konnten, basierend auf Befunden dieses Forschungsbereichs, konkrete Empfehlungen für die Fragebogengestaltung abgeleitet werden (z. B. Jobe & Loftus, 1991; Schwarz, 2007). Allerdings können kognitive Prozesse bislang lediglich mithilfe allgemeiner Modelle schematisch dargestellt werden (z. B. Krosnick, 1999), weshalb viele für die Fragebogenkonstruktion relevante Aspekte unklar bleiben. Beispielsweise *verstehen* Personen Begriffe weitestgehend gleich, jedoch weisen sie den Begriffen interindividuelle *Bedeutungen* zu, was sich wiederum auf die Itembeantwortung auswirken kann (Schwarz, 2007).

Fakt ist, dass bei der Fragebogenentwicklung eine Reihe von Aspekten zu berücksichtigen ist, beispielsweise bei der Definition des Messgegenstandes, der Wahl der Konstruktionsstrategie, der Festlegung von Testziel, Zielgruppe, Antwortformat und Testlänge sowie bei der Itemformulierung (z. B. Bühner, 2011; Jonkisz et al., 2012). Ein zufriedenstellendes Ergebnis ist nur dann zu erwarten, wenn dabei sowohl Erfahrung, Sprachgefühl und Intuition als auch wissenschaftliche Erkenntnisse über die einer Befragung zugrundeliegenden kognitiven und kommunikativen Prozesse berücksichtigt werden (Porst, 2014). Selbst für langjährige Fragebogen- und Umfrageforscher ist dies immer wieder eine Herausforderung: „Even after years of experience, no expert can write a perfect questionnaire“ (Bradburn et al., 2004, S. 317).

#### **4.1.3 Testtheoretische Überlegungen**

Mithilfe von Fragebögen sollen psychologische Konstrukte erfasst werden. Diese sind nicht direkt beobachtbar, weshalb sie als *latente Variablen* bezeichnet werden. Die Itemantworten sind das beobachtbare Verhalten einer Person, weshalb sie als *manifeste Variablen* bezeichnet werden. Die Items werden als *reflektive Indikatoren* für die latente Variable angesehen. Nur die Unterschiede in der Ausprägung auf der latenten Variablen sollen für die Zusammenhänge zwischen den manifesten Variablen ursächlich verantwortlich sein. Würde man die latente Dimension (statistisch) konstant halten, müssten die Zusammenhänge zwischen den manifesten Variablen verschwinden. Dies bedeutet, dass Variationen in der latenten Variablen Unterschiede im Antwortverhalten *aller* Items erklären, die Indikatoren für diese latente Variable sind. Die Ausprägung auf der latenten Variablen wird somit auf den Items *widergespiegelt* (Bühner, 2011).

Testtheorien beschäftigen sich mit dem Zusammenhang von zu erfassendem psychischem Merkmal und Testverhalten (Rost, 2004). Sie beruhen auf mathematisch-statistischen Modellen, die bestimmte Zusammenhänge zwischen Testwerten und den durch die Definition des Tests postulierten latenten Variablen erwarten lassen (Lienert & Raatz, 1998). Sie treffen Annahmen darüber, von welchen Komponenten die Itemantworten abhängig sind, beispielsweise von Personenfähigkeit und Itemschwierigkeit. Diese Modellannahmen können durch entsprechende Modelltests überprüft werden. Klassische Modelle nehmen kontinuierliche Daten an und werden unter der KTT zusammengefasst. Dazu gehören beispielsweise faktorenanalytische Modelle oder Strukturgleichungsmodelle

basierend auf Maximum-Likelihood-Schätzungen. In der Regel handelt es sich bei Fragebogendaten jedoch nicht um kontinuierliche, sondern um kategoriale Daten. Modelle, die diesen kategorialen Charakter mitberücksichtigen, werden innerhalb der Probabilistischen Testtheorie (PTT) subsummiert. Beispiele dafür sind das einparametrische logistische Modell (1PL-Modell) oder das zweiparametrische logistische Modell (2PL-Modell). Notwendige Voraussetzung für die Gültigkeit der meisten Testmodelle sind eindimensionale Messungen. Dies bedeutet, dass Zusammenhänge zwischen den Items nur von *einer* latenten Variablen kausal erklärt werden (Eid & Schmidt, 2014; Rost, 2004). Für eine ausführliche Darstellung des mathematisch-statistischen Hintergrundes der genannten Testtheorien wird auf weiterführende Literatur verwiesen (z. B. Bühner, 2011; Eid & Schmidt, 2014; Rost, 2004; van der Linden, 2016b).

Laut Rost (1999, S. 140) werden über 95 % der psychologischen Testverfahren auf Basis der KTT konstruiert. In Anlehnung an die KTT werden die Testgütekriterien Objektivität, Reliabilität und Validität definiert. Sie treffen Aussagen über die Qualität eines Tests (Bühner, 2011).

Unabhängig von den testtheoretischen Verfahren, mit denen Fragebogendaten analysiert werden, bleibt ein grundlegendes Problem: Zur Beurteilung der Güte eines Fragebogens werden Verfahren zur Überprüfung der Modellgültigkeit verwendet, die lediglich eine Aussage darüber treffen können, wie gut oder schlecht die jeweilige Modellpassung ist. Aussagen darüber, wie gut der Fragebogen das intendierte Konstrukt inhaltlich erfasst, sind dabei schwierig, u. a. deshalb, weil die wahren Werte dieser latenten Konstrukte immer unbekannt sind. Die Gültigkeit bzw. Validität eines Fragebogens wird oft mit Behelfslösungen und Korrelationen bestimmt, beispielsweise im Rahmen der Konstruktvalidität. Demnach wird ein Merkmal valide gemessen, wenn die Testergebnisse gering mit anderen abgrenzbaren Konstrukten korrelieren (divergente Validität) und/oder wenn die Testergebnisse hoch korrelieren mit denen eines anderen Messinstruments, welches das gleiche Konstrukt zu messen beansprucht (konvergente Validität) (Cronbach & Meehl, 1955). Problematisch ist, dass dabei zumeist wiederum Test- und Fragebogendaten herangezogen werden (Borsboom, Mellenbergh & van Heerden, 2004).

Die Existenz und Definition des Messgegenstandes wird vorausgesetzt, so wie ein Thermometer voraussetzt, dass die Temperatur ein bereits definierter physikalischer Begriff

ist. Im Falle der Temperaturmessung ist diese Annahme zutreffend, da Temperatur im Rahmen der Thermodynamik eindeutig definiert ist (Fischer, 1974). Im Gegensatz dazu sind die meisten psychologischen Konstrukte nicht vor und unabhängig von ihrer Messung eindeutig definiert. Für die meisten Konstrukte gibt es eine Vielzahl verschiedener Definitionen und Theorien (Kleinginna & Kleinginna, 1981). Die diesen Theorien zugrundeliegende Forschung basiert oft wiederum auf Test- und Fragebogendaten. Ist das Konstrukt durch die verwendeten Items bzw. Testdaten definiert, ist die Annahme kausaler Beziehungen zwischen Konstrukt und Items unzulässig. Das zugrundeliegende Merkmal darf nicht in einem Zirkelschluss als Ursache für das Testergebnis angesehen werden (Hartig, Frey & Jude, 2012).

#### **4.1.4 Rationale**

Unter Abschnitt 4.1.2 und 4.1.3 wurden die grundlegenden Herausforderungen bei der Test- und Fragebogenkonstruktion verdeutlicht. In Studie III wird der Prozess der Fragebogenkonstruktion für die physikalischen Merkmale Alter, Körpergröße und Körpergewicht durchlaufen. Sie sind die zu erfassenden „Konstrukte“ (latente Variablen), für die Items (reflektive Indikatoren) gefunden werden müssen. Die Messgrößen sind unabhängig von ihrer psychologischen Messung eindeutig definiert und ihre realen Werte sind bekannt bzw. der Fehler der physikalischen Messung vernachlässigbar. Mit diesem Vorgehen kann nicht nur die Modellpassung überprüft, sondern vor allem die Frage beantwortet werden, wie gut ein Fragebogen in der Lage ist, das zur Messung beabsichtigte Konstrukt zu erfassen.

Alle Teilnehmer sollen sich in den drei Merkmalsdimensionen Alter, Körpergröße und Körpergewicht mithilfe von Items selbst bewerten. Die Selbstbewertung wird entweder auf einer sechsstufigen (6PS) oder einer dichotomen Antwortskala (2PS) vorgenommen. In zwei Versionen des Fragebogens beantworten die Probanden dieselben Items mit Antwortskalen unterschiedlicher Breite. Anschließend werden die physikalischen Merkmale Alter, Körpergröße und Körpergewicht von allen Teilnehmern in offener Form selbst berichtet (z. B. gibt jeder Proband seine Körpergröße in Zentimetern an). Diese numerischen Werte stellen die realen Werte der untersuchten Merkmale dar.

Auf Basis vorhergehender Überlegungen ergeben sich folgende Forschungsfragen:

1. Wie sind die Skalen in Bezug auf Itemschwierigkeit und Itemtrennschärfe zu beurteilen?

2. Wie sind die Skalen bezüglich ihrer Reliabilität zu beurteilen?
3. Zeigen sich auffallende Korrelationen zwischen den Items (gewählte Antwortkategorie) und dem realen Wert einer Person sowie zwischen den Items unterschiedlicher Skalen?
4. Ergibt sich bei einer explorativen Faktorenanalyse eine dreifaktorielle Lösung, bei der alle Items auf den inhaltlich vorgesehenen Faktor laden?
5. Im Rahmen einer konfirmatorischen Faktorenanalyse werden zwei Modelle spezifiziert. Modell 1 nimmt drei miteinander korrelierende latente Faktoren (Alter, Körpergröße und Körpergewicht) an mit jeweils zwölf manifesten Variablen, die für jedes Merkmal theoretisch zugehörigen Items. Modell 2 ist ein hierarchisches Modell, in dem den drei latenten Faktoren (Alter, Körpergröße und Körpergewicht) ein Generalfaktor übergeordnet wird, der die Korrelationen zwischen den latenten Faktoren erklärt. Es soll überprüft werden, welches der beiden Modelle die Daten besser abbildet.
6. Werden im Rahmen probabilistischer Testmodelle die Daten besser durch ein 1PL-Modell oder ein 2PL-Modell (2PS) bzw. durch ein Partial-Credit-Modell oder ein Generalisiertes-Partial-Credit-Modell (6PS) beschrieben?
7. In welcher Höhe fallen die Determinationskoeffizienten  $R^2$  bei einfachen linearen Regressionen aus, mit Summenwerten, Faktorwerten oder Personenparametern als unabhängige Variablen und Alter, Körpergröße oder Körpergewicht bzw. BMI als abhängige Variablen?
8. Ergeben sich beim Merkmal Körpergewicht Unterschiede in den Ergebnissen, wenn die Regressionsanalysen auf Basis des BMI anstatt des absoluten Körpergewichts durchgeführt werden?
9. Ist das zwei- oder das sechsstufige Antwortformat besser geeignet, um die realen Werte in den Regressionsanalysen möglichst präzise vorherzusagen?

## **4.2 Methoden Studie III**

### **4.2.1 Präregistrierung**

Studie III wurde im Open Science Framework vor Beginn der Datenerhebung präregistriert. Unter folgendem Link ist die Präregistrierung einsehbar:

[https://osf.io/cz3uv/?view\\_only=9cac02db231b48629fea9ae53c3038b9](https://osf.io/cz3uv/?view_only=9cac02db231b48629fea9ae53c3038b9)

#### 4.2.2 Stichprobe

Die Gesamtstichprobe bestand aus 1 854 Teilnehmern im Alter von 18 bis 97 Jahren ( $M = 33$ ,  $SD = 16.0$ ), darunter 1 036 Frauen (55.9 %) und 818 Männer (44.1 %). Das Bildungsniveau verteilte sich wie folgt: 0.2 % Kein Schulabschluss, 4.2 % Hauptschulabschluss oder gleichwertig, 13.3 % Realschule oder gleichwertig, 47.6 % Abitur oder Fachabitur, 31.9 % Universitäts- oder Hochschulabschluss, 2.8 % Promotion oder Habilitation. 91.5 % der Befragten gaben Deutsch als Muttersprache an.

Die Gesamtstichprobe bestand aus zwei unabhängigen Substichproben, die sich auf die Bedingungen 2PS und 6PS verteilten. Tabelle 4.2.1. zeigt die Verteilung von Geschlecht, Bildungsniveau, Muttersprache und Alter innerhalb der zwei Bedingungen. Es ist ersichtlich, dass es sich um weitestgehend homogene Teilstichproben handelt. Aufgrund der linkssteilen Altersverteilung und Ausreißerwerten wird der Median anstatt des Mittelwerts als Maß der zentralen Tendenz berichtet.

Die Anwerbung der Probanden erfolgte in Zügen, Hochschulen, Universitäten, Volkshochschulen, Fitnessstudios sowie verschiedenen öffentlichen Plätzen in Deutschland und Österreich. Die Teilnahme war freiwillig und ohne Vergütung.

Tabelle 4.2.1.

*Verteilung von Geschlecht, Bildungsniveau, Muttersprache und Alter innerhalb der Bedingungen 2PS<sup>1</sup> und 6PS<sup>2</sup>*

Merkmal	2PS <sup>1</sup>		6PS <sup>2</sup>	
	<i>n</i>	%	<i>n</i>	%
Geschlecht				
Frauen	512	55.6	524	56.2
Männer	409	44.4	409	43.8
Bildungsniveau				
Kein Schulabschluss	1	0.1	2	0.2
Hauptschulabschluss oder gleichwertig	35	3.8	43	4.6
Realschule oder gleichwertig	129	14.0	119	12.8
Abitur oder Fachabitur	446	48.4	437	46.8
Universitäts- oder Hochschulabschluss	280	30.4	311	33.3
Promotion oder Habilitation	30	3.3	21	2.3
Muttersprache Deutsch				
Ja	837	90.9	860	92.2
Nein	84	9.1	73	7.8
	<i>Md</i>	<i>SD</i>	<i>Md</i>	<i>SD</i>
Alter	25	15.9	25	16.1
Gesamtteilnehmerzahl	921		933	

*Anmerkungen.*

<sup>1</sup>Dichotome Antwortskala.

<sup>2</sup>Sechsstufige Antwortskala.

*n* = Anzahl; % = prozentualer Anteil innerhalb der Teilstichprobe; *Md* = Median; *SD* = Standardabweichung.

### 4.2.3 Skalenkonstruktion

In Studie III wurden für die Merkmale Alter, Körpergröße und Körpergewicht Items konstruiert. Den Empfehlungen zur Itemkonstruktion folgend wurden zwei Konstruktionsstrategien miteinander kombiniert (z. B. Bühner, 2011; Jonkisz et al., 2012): Der deduktive Testkonstruktionsansatz und der Prototypenansatz.

Im ersten Konstruktionsschritt wurden die zu untersuchenden Merkmale definiert und spezifiziert. Alter wurde definiert als chronologisches Alter bzw. Lebensalter entsprechend der vergangenen Zeit in Jahren seit der Geburt einer Person (Montepare & Lachman, 1989; Schwall, 2012). Körpergröße wurde definiert als Größe einer aufrecht stehenden Person von der Fußsohle bis zum Scheitel in Zentimetern und Körpergewicht als die physikalische Masse



einer Person in Kilogramm (Martin, 1929). Als Zielgruppe wurden Frauen und Männer ab 18 Jahren über alle Bildungsniveaus festgelegt.

Abhängig vom Geltungsbereich des zu erfassenden Merkmals können unterschiedliche Testlängen angemessen sein. So lassen sich homogene Merkmalsdimensionen mit weniger Items erfassen als heterogene Merkmalsdimensionen (Jonkisz et al., 2012). Für Kurzskalen werden mindestens drei Items pro Dimension empfohlen (Lang & Lüdtke, 2005). In Studie III soll weder eine Kurzskala konstruiert werden, wie beispielsweise das *Big-Five-Inventory-Shortversion* mit drei Items je Skala (Gerlitz & Schupp, 2005), noch ein ausgereiftes Inventar, wie beispielsweise das *NEO-PI-R* mit 48 Items je Hauptskala (Ostendorf & Angleitner, 2004). Deshalb wurde, in Anlehnung an den *NEO-FFI* (Borkenau & Ostendorf, 2008), die Testlänge auf zwölf Items je Merkmal festgelegt. Als Kurzversion des *NEO-PI-R* weist der *NEO-FFI* mit zwölf Items je Skala zufriedenstellende psychometrische Kennwerte auf (Borkenau & Ostendorf, 2008; Schmitz, Hartkamp, Baldini, Rollnik & Tress, 2001).

*Deduktiver Testkonstruktionsansatz.* Voraussetzung für die Testkonstruktion ist zunächst differenziertes Wissen über die Unterschiedlichkeit von Personen hinsichtlich des interessierenden Merkmals (Bühner, 2011). Zur Gewinnung dieser Informationen wurden Experteninterviews geführt (vgl. Bogner, Littig & Menz, 2014). Als Experten herangezogen wurden Vertreter von Merkmalsextreemen, d. h. Personen, die in den relevanten Merkmalen mindestens eine Standardabweichung vom Populationsmittelwert abweichen. Als Populationskennwerte wurden für Deutschland repräsentative Daten des GESIS - Leibniz-Institut für Sozialwissenschaften (2015) herangezogen (Alter in Jahren Frauen:  $M = 49$ ,  $SD = 17.3$ ; Alter in Jahren Männer:  $M = 50$ ,  $SD = 17.7$ ; Körpergröße in Zentimetern Frauen:  $M = 166$ ,  $SD = 6.5$ ; Körpergröße in Zentimetern Männer:  $M = 179$ ,  $SD = 7.2$ ; Körpergewicht in Kilogramm Frauen:  $M = 70$ ,  $SD = 14.4$ ; Körpergewicht in Kilogramm Männer:  $M = 85$ ,  $SD = 15.1$ ). Für jedes Merkmalsextreem wurden zwei Männer und zwei Frauen interviewt. Das Alter lag bei 21 (zwei Mal), 67 und 76 Jahren (Frauen) bzw. bei 19 (zwei Mal), 73 und 77 Jahren (Männer), die Körpergröße lag bei 155, 158, 183 und 187 Zentimetern (Frauen) bzw. bei 170, 171, 194 und 203 Zentimetern (Männer) und das Körpergewicht lag bei 50, 55, 90 und 105 Kilogramm (Frauen) bzw. bei 53, 69, 119 und 135 Kilogramm (Männer). Bei acht Interviews je Merkmal ergaben sich 24 Interviews insgesamt. Die Dauer pro Interview lag zwischen 14 und 28 Minuten. Alle Interviews wurden aufgezeichnet und transkribiert. Die

vollständigen Transkripte wurden aufgrund datenschutzrechtlicher Bestimmungen für personenbezogene Daten nicht veröffentlicht (vgl. EU Datenschutz-Grundverordnung). Anschließend wurden alle Interviewaussagen thematisch sortiert. Auf Basis der identifizierten Themenblöcke wurden, unter Berücksichtigung der Richtlinien zur Itemformulierung, 79 Items formuliert.

*Prototypenansatz.* Der Prototypenansatz nutzt die Idee, dass Menschen für jede Eigenschaft eine prototypische und meist ähnliche Vorstellung von merkmalspezifischen Verhaltensweisen haben (Broughton, 1984; Cantor & Mischel, 1977; Rosch, 1973). Zur Sammlung dieser prototypischen Vorstellungen wurde ein Fragebogen mit offenem Antwortformat konstruiert. Der Fragebogen findet sich in Anhang C (Abbildungen C.1. bis C.4.). Darin wurden die Befragten gebeten, sich eine ihnen bekannte Person vorzustellen, die ein typischer Vertreter eines bestimmten Merkmals ist (z. B. eine sehr große Person). Sodann sollten sie typische Verhaltensweisen aufschreiben, die auf dieses Merkmal zurückzuführen sind. Bei den Befragten handelte es sich um 123 Psychologiestudenten der LMU (Bachelor und Master) im Alter von 19 bis 49 Jahren ( $M = 25$ ,  $SD = 6.6$ ), darunter 99 Frauen (80.5 %) und 24 Männer (19.5 %). Bis auf eine Person gaben alle Personen an, dass sie Abitur haben. Die Teilnahme war freiwillig und ohne Vergütung. Anschließend wurden alle Merkmalsbeschreibungen thematisch sortiert. Auf Basis der identifizierten Themenblöcke wurden, unter Berücksichtigung der Richtlinien zur Itemformulierung, 59 Items formuliert.

*Phase I Itemanalyse.* Die aus beiden Ansätzen hervorgegangenen Items wurden einem kognitiven Pretest unterzogen (vgl. Lenzner, Neuert & Otto, 2015). Mithilfe kognitiver Interviews wurde u. a. überprüft, wie Itemaussagen und Begriffe interpretiert werden und ob diese mit konstruktfernden Vorstellungen verknüpft sind. Dieses Verfahren stellt sicher, dass jedes Item das beabsichtigte Konstrukt erfasst. Missverständliche Items werden dabei umformuliert oder aussortiert (Faulbaum et al., 2009). Angewendet wurden die kognitiven Verfahren *probing*, *paraphrasing*, *concurrent-think-aloud* und *retrospektive-think-aloud* (Prüfer & Rexroth, 1996; Prüfer & Rexroth, 2005). Im Rahmen solcher Pretests werden zwischen fünf und 30 Interviews empfohlen, wobei die gravierendsten Probleme bereits bei geringer Interviewanzahl identifiziert werden können (Willis, 2005). Aufgrund begrenzter Zeit und Ressourcen wurden in dieser Studie acht Interviews durchgeführt. Interviewt

wurden Personen aus der späteren Zielgruppe (vier Frauen und vier Männer) im Alter von 21 bis 77 Jahren ( $M = 44$ ,  $SD = 20.8$ ). Das Bildungsniveau erstreckte sich von Hauptschulabschluss bis Universitätsabschluss. Die Dauer je Interview lag zwischen zwei und dreieinhalb Stunden. Alle Interviews wurden von der Autorin durchgeführt. Die Teilnahme war freiwillig und ohne Vergütung. Die vollständigen Dokumentationen beider Itemkonstruktionsprozesse (Interviewaussagen bzw. Merkmalsbeschreibungen aus dem deduktiven Konstruktionsansatz bzw. aus dem Prototypenansatz, Itementwurf, schriftliche Dokumentation der kognitiven Interviews und der daraus abgeleiteten Verwerfungen bzw. Umformulierungen der Items) sind unter folgendem Link im Open Science Framework einsehbar: [https://osf.io/qvnjr/?view\\_only=9cac02db231b48629fea9ae53c3038b9](https://osf.io/qvnjr/?view_only=9cac02db231b48629fea9ae53c3038b9)

*Phase II Itemanalyse.* Nach dem kognitiven Pretest verblieben 61 Items. Diese wurden einer größeren Stichprobe in Form eines Online-Pilottests mit dem Ziel vorgelegt, eine endgültige Itemauswahl für die Haupterhebung zu treffen. Auf Basis des Pilottests wurden pro Merkmal zwölf Items ausgewählt, die für die zu messenden Merkmale psychometrisch und inhaltlich betrachtet am geeignetsten erschienen. Es wurde ein sechstufiges Skalenformat verwendet, bei dem die verbale und numerische Verankerung analog zur Haupterhebung erfolgte. Für die Befragung wurde ein Onlinefragebogen mit der Befragungssoftware SoSci Survey (Leiner, 2014) erstellt und den Teilnehmern auf [www.soscisurvey.de](http://www.soscisurvey.de) zugänglich gemacht. Eine Übersicht aller Pilottest-Items findet sich in Anhang C (Tabelle C.1.). Ein exemplarischer Auszug des Onlinefragebogens findet sich als Screenshot in Anhang C (Abbildung C.5.).

Der Pilottest wurde im Zeitraum vom 05. bis 08. November 2017 durchgeführt. Insgesamt wurden die Daten von 456 Teilnehmern gesammelt. Davon wurden 66 Probanden ausgeschlossen, weil sie den Fragebogen nicht vollständig bearbeitet hatten. Somit ergab sich eine Stichprobe von 390 Teilnehmern im Alter von 18 bis 77 Jahren ( $M = 32$ ,  $SD = 12.8$ ), darunter 302 Frauen (77.4 %) und 88 Männer (22.6 %). Das Bildungsniveau verteilte sich wie folgt: 0 % Kein Schulabschluss, 3.3 % Hauptschulabschluss/Volksschule oder gleichwertig, 17.2 % Realschule oder gleichwertig, 35.9 % Fachabitur oder Abitur, 40.8 % Hochschulabschluss oder Universitätsabschluss, 2.8 % Promotion oder Habilitation. 95.4 % der Befragten gaben Deutsch als Muttersprache an. Die Anwerbung der Probanden erfolgte über das soziale Netzwerk Facebook, über private

E-Mail-Verteiler der Autorin sowie über den Studien-Informationen-Service der LMU. Die Teilnahme war freiwillig und ohne Vergütung.

Als Reliabilitätsmaßzahlen wurden Cronbachs Alpha (Cronbach, 1951) und McDonalds Omega (McDonald, 1970) berichtet. Eine Abgrenzung beider Maßzahlen wird in Abschnitt 4.2.6 vorgenommen. Bei der Zuordnung der Items zu den Skalen wurde der theoretisch vorgesehenen Einteilung gefolgt. Die Itemselektion für den endgültigen Fragebogen basierte auf Kennwerten der KTT und Ergebnissen einer explorativen Faktorenanalyse (EFA) mit drei Faktoren (Maximum Likelihood mit Promax als obliques Rotationsverfahren). Für die Evaluation nach der KTT wurden die part-whole-korrigierte Trennschärfe und die Schwierigkeit für jedes Item ermittelt. Als Schwierigkeitsindex wurde der Mittelwert der Itemantworten über alle Personen herangezogen (Bühner, 2011). Negativ gepolte Items (13p, 14p, 21p, 35p, 36p, 37p, 38p, 39p, 40p, 41p, 43p, 52p, 53p, 54p, 55p, 56p, 57p und 58p) wurden vor den Analysen entsprechend der vorgesehenen Schlüsselrichtung des zu erfassenden Merkmals umgepolt.

Im ersten Selektionsschritt wurden die Items für die endgültige Testversion ausgeschlossen, deren Hauptladung nicht auf den vorgesehenen Faktor fällt, deren Haupt- und Nebenladung nahezu gleich ausfallen oder die auf keinen Faktor laden (Alter: 20p; Körpergewicht: 44p, 55p). Von den verbliebenen Items wurden sodann die Items mit Trennschärfen  $\leq .30$  für die endgültige Testversion ausgeschlossen (Alter: 7p, 11p, 13p, 18p, 19p, 21p). Nach dem ersten Selektionsschritt wurden somit neun Items ausgeschlossen, die in den folgenden Selektionsschritten nicht mehr berücksichtigt wurden. Im zweiten Selektionsschritt wurde bei inhaltsähnlichen Items das Item mit der höheren Ladung und Trennschärfe ausgewählt (ausgewähltes Item an erster Stelle; Alter: 6p vs. 16p, 3p vs. 4p; Körpergröße: 26p vs. 38p, 28p vs. 27p vs. 40p, 32p vs. 35p, 33p vs. 36p, 41p vs. 34p, 39p vs. 23p, 42p vs. 43p; Körpergewicht: 49p vs. 56p, 50p vs. 57p, 54p vs. 45p, 59p vs. 58p). Im dritten Selektionsschritt (nur bei Körpergröße relevant) wurden von den nach Selektionsschritt zwei noch verbliebenen Items diejenigen mit den niedrigsten Ladungen und Trennschärfen ausgeschlossen (29p, 30p). Da bei den Merkmalen Alter und Körpergewicht nach Selektionsschritt zwei weniger als zwölf Items ausgewählt worden wären, wurden inhaltsähnliche Items beibehalten (Alter: 9p und 14p; Körpergewicht: 48p und 52p) und auf Selektionsschritt drei verzichtet. Nach Abschluss der Itemselektion wurden

die Items 32p und 52p umformuliert. Ziel war eine sprachliche Vereinfachung (32p) bzw. die Auflösung der sprachlichen Ähnlichkeit zum inhaltsähnlichen Item (48p und 52p). Die Ergebnisse der Itemanalysen sind in Anhang C (Tabellen C.2., C.3., C.4. und C.5.) einzusehen. Eine Übersicht aller Items für die Haupterhebung findet sich in Anhang C (Tabelle C.6.). Verweise auf Items im Ergebnisteil verwenden die in Tabelle C.6. eingeführte Nummerierung.

#### **4.2.4 Material**

Zur Untersuchung der Fragestellung wurde ein Fragebogen erstellt, in dem den Probanden zunächst 36 Itemaussagen (zwölf Items je Merkmal) vorgegeben wurden. Bei jeder Aussage sollten sie markieren, inwieweit sie dieser zustimmen. Unter zwei Bedingungen ohne Messwiederholung konnte dies entweder auf einer dichotomen Skala (2PS) oder einer sechsstufigen Skala (6PS) erfolgen. Bei jeder Skala wurden die Skalenpole mit den Worten *trifft nicht zu* und *trifft zu* beschriftet. Es erfolgte keine weitere verbale Verankerung bei der sechsstufigen Skala. Die numerische Verankerung war null und eins (2PS) bzw. eins bis sechs (6PS). Zum besseren Verständnis der Skalen wurden am Anfang des Fragebogens entsprechende Bearbeitungshinweise gegeben. Die gewählten Antwortkategorien gingen als die beobachteten Werte einer Person in die Datenanalyse ein.

Am Ende des Fragebogens wurden Alter, Körpergröße und Körpergewicht der Teilnehmer abgefragt. Sie bildeten die realen Werte der untersuchten Merkmale ab. Analog zu den Studien I und II sind davon die wahren Werte eines Merkmals im Sinne der KTT abzugrenzen (vgl. Abschnitt 2.2.2). Auch in dieser Studie wird sich nicht auf die Definition der KTT bezogen, weshalb die sprachliche Abgrenzung beibehalten wurde. Zudem wurden mithilfe vorgegebener Auswahlmöglichkeiten Geschlecht (*weiblich* oder *männlich*), Muttersprache (Deutsch *ja* oder *nein*) und höchster Bildungsabschluss (*Kein Schulabschluss*, *Hauptschulabschluss oder gleichwertig*, *Realschule oder gleichwertig*, *Abitur oder Fachabitur*, *Universitäts- oder Hochschulabschluss*, *Promotion oder Habilitation*) erfasst. Beide Versionen des Fragebogens finden sich in Anhang C (Abbildungen C.6. bis C.10.).

#### **4.2.5 Erhebung**

Die Datenerhebung erfolgte vom 16. November bis 28. Dezember 2017 in Form eines Papier-Bleistift-Fragebogens. Die Papier-Bleistift-Erhebung wurde anstatt einer Online-Erhebung gewählt, da sich die Online-Rekrutierung großer und unbezahlter Stichproben als

zunehmend schwierig herausstellte (z. B. Kennzeichnung als Spam in sozialen Netzwerken). Voraussetzung für die Teilnahme war ein Mindestalter von 18 Jahren, da es sich um eine Folgestudie der Studien I und II handelte und die Stichproben vergleichbar sein sollten. Darüber hinaus gab es keine Ein- bzw. Ausschlusskriterien.

Zunächst wurden alle Teilnehmer mithilfe einer separaten Textseite über die Freiwilligkeit und Anonymität der Teilnahme aufgeklärt sowie über den Zweck der Untersuchung und die zu erhebenden Daten informiert. Sehr junge Teilnehmer wurden nach ihrem Alter gefragt, um sicherzustellen, dass nur Teilnehmer über 18 Jahren rekrutiert wurden. Waren die Befragten mit den Durchführungsbedingungen einverstanden, wurde ihnen ein Fragebogen zur Bearbeitung ausgehändigt. Die Textseite konnten die Teilnehmer behalten. Um eine zufallsgesteuerte Zuweisung der Teilnehmer zu einer der zwei Bedingungen (2PS oder 6PS) ohne Messwiederholung annähernd zu gewährleisten, waren die Fragebogenversionen abwechselnd sortiert, sodass alternierend eine der beiden Versionen ausgehändigt wurde. Nach Rückgabe des ausgefüllten Fragebogens durch den Probanden wurde dieser auf Vollständigkeit überprüft und sich für die Teilnahme bedankt.

#### **4.2.6 Analysen**

Die Datenanalyse erfolgte mit der Open Source Statistiksoftware R (R Core Team, 2017). Für die Berechnung wurden die R-Pakete „foreign“ (R Core Team, 2016), „dplyr“ (Wickham & Francois, 2016), „psych“ (Revelle, 2017), „lavaan“ (Rosseel, 2012), „semPlot“ (Epskamp & Stuber, 2017), „MBESS“ (Kelley, 2018), „GPArotation“ (Bernaards & Jennrich, 2005) und „ltm“ (Rizopoulos, 2006) verwendet. Die Diagramme wurden mit Microsoft PowerPoint oder Microsoft Excel erstellt.

Beim verwendeten Fragebogen gaben die Teilnehmer entweder auf einer dichotomen oder sechsstufigen Antwortskala den Grad der Zustimmung zu den jeweiligen Aussagen an. Die Antwortalternativen wurden wie folgt kodiert: Bei der dichotomen Antwortskala *trifft nicht zu* mit der Ziffer 0 und *trifft zu* mit der Ziffer 1; bei der sechsstufigen Antwortskala *trifft nicht zu* mit der Ziffer 1, *trifft zu* mit der Ziffer 6 und die dazwischenliegenden Kategorien aufsteigend mit den Ziffern 2, 3, 4 und 5. Negativ gepolte Items (10, 21, 22, 23, 31, 32 und 33) wurden vor den Analysen entsprechend der vorgesehenen Schlüsselrichtung des zu erfassenden Merkmals umgepolt. Die deskriptive Datenanalyse und die Korrelationsanalyse wurden getrennt für Männer und Frauen durchgeführt. Alle weiteren Analysen wurden für

beide Geschlechter gemeinsam durchgeführt, obwohl geschlechtsspezifische Unterschiede nicht ausgeschlossen werden konnten. Getrennte Datenauswertungen hätten zu Substichprobengrößen geführt, die für die geplanten Analysen unbefriedigend gewesen wären. Deshalb wurden mögliche unentdeckte Geschlechtsunterschiede den größeren Gesamtstichproben untergeordnet.

*Deskriptive Statistik.* Bei der deskriptiv-statistischen Datenanalyse wurden Mittelwert, Standardabweichung, Median, kleinster Wert und größter Wert der realen Werte für jede Bedingung und für jedes Merkmal jeweils für Frauen und Männer tabellarisch gegenübergestellt. Zu den Häufigkeiten der realen Werte wurde für jede Bedingung und für jedes Merkmal ein Histogramm erstellt. Darin wurden beide Geschlechter getrennt dargestellt. Zudem wurden Mittelwert, Median und Standardabweichung für die drei erhobenen Merkmale innerhalb beider Bedingungen getrennt nach Bildungsniveau tabellarisch gegenübergestellt. Da einige Bildungsabschlüsse nur selten beobachtet wurden (z. B. *Kein Schulabschluss*), wurden zur einfacheren Interpretation die Kategorien *Kein Schulabschluss*, *Hauptschulabschluss oder gleichwertig* und *Realschule oder gleichwertig* zur Kategorie *Kein Abitur* zusammengefasst. Die Kategorie *Abitur oder Fachabitur* blieb inhaltlich unverändert, wurde aber in *Abitur* umbenannt. Die Kategorien *Universitäts- oder Hochschulabschluss* und *Promotion oder Habilitation* wurden zur Kategorie *Hochschulabschluss* zusammengefasst.

*Korrelations- und Reliabilitätsanalyse.* Zur Untersuchung von Zusammenhängen zwischen gewählter Antwortkategorie und realem Wert einer Person sowie von Zusammenhängen zwischen den Items wurden getrennt für beide Geschlechter Spearman-Korrelationen (6PS) bzw. punktbiseriale Korrelationen (2PS) berechnet. Dafür wurden Korrelationsmatrizen mit den realen Werten (Alter, Körpergröße und Körpergewicht) und allen Items erstellt. Es handelt sich dabei um ergänzende explorative Analysen, die über die präregistrierten Analysen hinaus gehen.

Zur Reliabilitätsschätzung der drei Merkmalsdimensionen wurden beim polytomen Antwortformat Cronbachs Alpha (Cronbach, 1951) und beim dichotomen Antwortformat Kuder-Richardson-20 (Kuder & Richardson, 1937) herangezogen. Bei der Zuordnung der Items zu den Skalen wurde der theoretisch vorgesehenen Einteilung gefolgt. Als weiteres Reliabilitätsmaß wurde, nach der Empfehlung von Zinbarg, Revelle, Yovel und Li (2005),

McDonalds Omega (McDonald, 1970) berichtet. Die Grundlage für eine präzise Schätzung der Reliabilität ist die Geltung eines testtheoretischen Messmodells (Bühner, 2011). Cronbachs Alpha bzw. Kuder-Richardson-20 sind nur dann angemessene Maße für die Reliabilität des Summenwerts, wenn ein mindestens essentiell tau-äquivalentes Messmodell vorliegt. Andernfalls handelt es sich lediglich um eine Mindestschätzung der Reliabilität (Miller, 1995; Raykov, 1997). In vielen Fällen liegen höchstens tau-kongenerische Messmodelle vor (Cho, 2016), bei denen McDonalds Omega das angemessene Maß für die Reliabilität des Summenwerts ist (Sočan, 2000; Zinbarg et al., 2005).

*Itemanalyse.* Für jedes Item wurde der Mittelwert der Itemantworten (6PS) bzw. der relative Anteil an *trifft zu*-Antworten (2PS) über alle Personen als Schwierigkeitsindex ermittelt. In der Testkonstruktion wird in der Regel eine breite Streuung der Schwierigkeitsindizes angestrebt, um in allen Bereichen der Merkmalsausprägung gut zwischen Personen differenzieren zu können (Bühner, 2011). Zur Feststellung der Differenzierungsfähigkeit der Items wurde für jedes Item die Standardabweichung ermittelt. Für die polytomen Items (6PS) wurden zusätzlich Histogramme erstellt. Als Trennschärfekoeffizient wurde die part-whole-korrigierte Trennschärfe ermittelt. Weisen Items sehr ähnliche Trennschärfen auf, deutet dies deskriptiv auf die Passung eines tau-äquivalenten Messmodells hin (Bühner, 2011; Eid & Schmidt, 2014).

*Explorative Faktorenanalyse.* Bei der Datenanalyse im Rahmen der EFA wurden als Verfahren zur Bestimmung der optimalen Faktorenanzahl Parallelanalysen (Horn, 1965) mit 500 Iterationen durchgeführt. Für beide Bedingungen wurde eine EFA mit der sich aus der Parallelanalyse ergebenden Faktorenanzahl durchgeführt und eine EFA mit den drei theoretisch angenommenen Faktoren (Alter, Körpergröße und Körpergewicht). Für die EFA wurden die Mustermatrizen der tetrachorischen (2PS) bzw. polychorischen Korrelationen (6PS) analysiert (Eid, Gollwitzer & Schmitt, 2015). Als Schätzmethode wurde die Maximum Likelihood Methode und als obliques Rotationsverfahren Promax herangezogen. Es wurde überprüft, ob die von der Parallelanalyse vorgeschlagene Faktorenlösung empirisch interpretierbar ist. Bei der dreifaktoriellen Lösung wurde untersucht, welche Items auf welchen Faktor laden und in welcher Höhe die Nebenladungen auf andere Faktoren ausfallen.



*Konfirmatorische Faktorenanalyse.* Im Rahmen der konfirmatorischen Faktorenanalyse (CFA) wurden zwei Modelle spezifiziert und miteinander verglichen. Modell 1 geht von drei miteinander korrelierenden latenten Faktoren aus, mit jeweils zwölf manifesten Variablen, den für dieses Merkmal theoretisch zugehörigen Items. Jedes Item lädt auf genau einem Faktor (zwölf pro Faktor). Die Varianz der latenten Faktoren wurde jeweils auf Eins fixiert. Modell 2 ist ein hierarchisches Modell. Es nimmt identisch zu Modell 1 drei latente Faktoren (Alter, Körpergröße und Körpergewicht) an, mit jeweils zwölf manifesten Variablen, den diesem Merkmal theoretisch zugehörigen Items. Jedes Item lädt auf genau einem Faktor (zwölf pro Faktor). Ergänzend wird den drei latenten Faktoren ein latenter Generalfaktor übergeordnet, der die Korrelationen zwischen den drei Faktoren erklärt. Bei allen drei Faktoren (Alter, Körpergröße und Körpergewicht) wurde jeweils die Ladung des ersten Items auf Eins fixiert. Die Varianz des Generalfaktors wurde ebenfalls auf Eins fixiert. Die Parameterschätzung erfolgte für beide Modelle nach der WLSMV (weighted least squares means and variance adjusted) Methode (Muthén & Muthén, 1998–2017). Zur Beurteilung der Modellpassung wurde das Ergebnis des Chi-Quadrat-Modell-Tests herangezogen. Als relativer Modellfit-Index wurde der CFI (comparative fit index) und als absolute Modellfit-Indizes der RMSEA (root mean square error of approximation) sowie der SRMR (standardized root mean residual) herangezogen. Den Empfehlungen von Hu und Bentler (1999) sowie Beauducel und Wittmann (2005) folgend, wurden als Grenzwerte für den CFI  $> .95$ , für den RMSEA  $\leq .06$  (da  $n > 250$ ) und für den SRMR  $\leq .11$  festgelegt.

*Probabilistische Testmodelle.* Die Datenanalyse erfolgte zusätzlich mithilfe von probabilistischen Testmodellen aus der Item Response Theorie. Unter Bedingung 2PS wurde ein 1PL-Modell einem 2PL-Modell gegenübergestellt, unter Bedingung 6PS wurde ein Partial-Credit-Modell (PCM) einem Generalisierten-Partial-Credit-Modell (GPCM) gegenübergestellt. Es wurde jeweils überprüft, welches Modell die Daten besser beschreibt. Alle Modelle wurden mit der Marginal Maximum Likelihood Methode geschätzt. Zur Normierung wurde in allen Modellen eine Normalverteilung mit Erwartungswert Null für die latente Variable angenommen. Im 1PL-Modell und PCM wurde die Varianz der latenten Variablen geschätzt. Im 2PL-Modell und GPCM musste die Varianz zur Normierung auf Eins fixiert werden. Zur Beurteilung der besten Modellpassung wurde als absoluter Modelltest der Pearson Chi-Quadrat-Test durchgeführt (Eid & Schmidt, 2014). Der  $p$ -Wert wurde

sowohl basierend auf der approximativen Chi-Quadrat-Verteilung der Teststatistik als auch mithilfe eines parametrischen Bootstrap-Verfahrens mit 300 Iterationen bestimmt (Langeheine, Pannekoek & van de Pol, 1996). Neben dem absoluten Modellfit wurden die unterschiedlich flexiblen Modelle jeweils mithilfe des Likelihood Ratio Tests verglichen. Außerdem wurden zum Modellvergleich die informationstheoretischen Maße AIC (Akaike Information Criterion) und BIC (Bayes Information Criterion) herangezogen (Rost, 2004). Beim Merkmal Körpergewicht wurde beim GPCM die Anzahl der Iterationen des Schätzalgorithmus auf 300 erhöht, um die Konvergenz des Modells zu gewährleisten.

*Regressionsanalyse.* Abschließend wurden einfache lineare Regressionen durchgeführt, mit Summenwerten, Faktorwerten (für das CFA-Modell mit dem relativ besseren Modellfit) oder Personenparametern (für das probabilistische Testmodell mit dem relativ besseren Modellfit) als unabhängige Variablen und Alter, Körpergröße oder Körpergewicht bzw. BMI als abhängige Variablen. In der CFA wurden Bartlett Faktorwerte verwendet. In den probabilistischen Testmodellen wurden für die Personenparameter expected a posteriori Schätzer verwendet. Berichtet wurden die Determinationskoeffizienten  $R^2$ .

Das Signifikanzniveau wurde bei allen Tests auf  $\alpha = 5\%$  festgelegt, falls nicht anders angegeben.

### **4.3 Ergebnisse Studie III**

#### **4.3.1 Deskriptive Statistik**

Die Tabellen 4.3.1. und 4.3.2. zeigen Mittelwert, Standardabweichung, Median, kleinsten Wert und größten Wert der realen Werte für alle Merkmale und Bedingungen getrennt nach Geschlecht. Dabei ist ersichtlich, dass die Lagemaße der zentralen Tendenz bei den körperbezogenen Merkmalen bei Frauen geringer ausfallen als bei Männern. Beim Merkmal Alter fallen diese für beide Geschlechter annähernd gleich aus. Abgesehen von den kleinsten und größten Werten bei den Merkmalen Körpergröße und Körpergewicht stimmen die Streuungsmaße für beide Geschlechter in etwa überein.

Tabelle 4.3.1.

Mittelwert (*M*), Standardabweichung (*SD*), Median (*Md*), Minimum (*Min*) und Maximum (*Max*) reale Werte getrennt für Frauen und Männer – Bedingung 2PS<sup>1</sup>

Merkmal	Frauen					Männer				
	<i>M</i>	<i>SD</i>	<i>Md</i>	<i>Min</i>	<i>Max</i>	<i>M</i>	<i>SD</i>	<i>Md</i>	<i>Min</i>	<i>Max</i>
Alter	32.9	15.4	25	18	89	32.6	16.5	25	18	84
Körpergröße	168	6.4	168	149	190	181	7.2	180	158	205
Körpergewicht	62.4	12.0	60	39	130	79.9	12.5	78	51	123

Anmerkungen.

*M* = Mittelwert; *SD* = Standardabweichung; *Md* = Median; *Min* = kleinster Wert; *Max* = größter Wert.

<sup>1</sup> Dichotome Antwortskala.

Tabelle 4.3.2.

Mittelwert (*M*), Standardabweichung (*SD*), Median (*Md*), Minimum (*Min*) und Maximum (*Max*) reale Werte getrennt für Frauen und Männer – Bedingung 6PS<sup>1</sup>

Merkmal	Frauen					Männer				
	<i>M</i>	<i>SD</i>	<i>Md</i>	<i>Min</i>	<i>Max</i>	<i>M</i>	<i>SD</i>	<i>Md</i>	<i>Min</i>	<i>Max</i>
Alter	32.7	15.7	25	18	97	32.4	16.7	25	18	90
Körpergröße	167	6.6	167	150	199	181	7.1	180	160	200
Körpergewicht	63.5	12.7	60	40	123	79.6	14.0	78	50	178

Anmerkungen.

*M* = Mittelwert; *SD* = Standardabweichung; *Md* = Median; *Min* = kleinster Wert; *Max* = größter Wert.

<sup>1</sup> Sechsstufige Antwortskala.

Die Histogramme der Häufigkeiten der realen Werte für die drei erhobenen Merkmale spiegeln die deskriptiven Maße aus den Tabellen 4.3.1. und 4.3.2. wider. Bei den Merkmalen Körpergewicht und Körpergröße lassen sich bimodale Verteilungen erkennen. Beim Merkmal Alter zeigt sich eine unimodale linkssteile Verteilung. Die Histogramme finden sich in Anhang C (Abbildungen C.11., C.12. und C.13.).

Die Tabellen 4.3.3. und 4.3.4. zeigen Mittelwert, Standardabweichung und Median für alle Merkmale und Bedingungen getrennt nach Bildungsniveau und Geschlecht. Dabei ist ersichtlich, dass die Lagemaße der zentralen Tendenz beim Merkmal Alter über alle Bedingungen und für beide Geschlechter beim Bildungsniveau *Abitur* am geringsten ausfallen. Beim Merkmal Körpergewicht fallen diese über alle Bedingungen und für beide

Geschlechter beim Bildungsniveau *Kein Abitur* am höchsten aus. Beim Merkmal Körpergröße zeigen sich keine auffälligen Unterschiede in den Lagemaßen der zentralen Tendenz zwischen den Bedingungen.

Tabelle 4.3.3.

Mittelwert (*M*), Standardabweichung (*SD*) und Median (*Md*) der realen Werte getrennt nach Bildungsniveau und Geschlecht – Bedingung 2PS<sup>1</sup>

Merkmal	Kein Abitur				Abitur				Hochschulabschluss			
	<i>n</i>	<i>M</i>	<i>SD</i>	<i>Md</i>	<i>n</i>	<i>M</i>	<i>SD</i>	<i>Md</i>	<i>n</i>	<i>M</i>	<i>SD</i>	<i>Md</i>
Alter												
Frauen	99	47	15.5	48	239	25	11.4	20	174	36	13.1	31
Männer	66	46	18.3	49	207	25	11.3	22	136	37	16.7	29
Körpergröße												
Frauen	99	167	6.8	167	239	168	6.1	168	174	167	6.5	168
Männer	66	179	7.3	179	207	182	7.0	182	136	182	7.1	180
Körpergewicht												
Frauen	99	68	14.0	64	239	60	9.8	58	174	63	12.5	60
Männer	66	83	13.7	82	207	78	12.6	76	136	82	11.4	80

Anmerkungen.

*n* = Anzahl; *M* = Mittelwert; *SD* = Standardabweichung; *Md* = Median.

<sup>1</sup> Dichotome Antwortskala.

Tabelle 4.3.4.

Mittelwert (*M*), Standardabweichung (*SD*) und Median (*Md*) der realen Werte getrennt nach Bildungsniveau und Geschlecht – Bedingung 6PS<sup>1</sup>

Merkmal	Kein Abitur				Abitur				Hochschulabschluss			
	<i>n</i>	<i>M</i>	<i>SD</i>	<i>Md</i>	<i>n</i>	<i>M</i>	<i>SD</i>	<i>Md</i>	<i>n</i>	<i>M</i>	<i>SD</i>	<i>Md</i>
Alter												
Frauen	92	47	17.0	47.5	251	25	11.5	21	181	35	13.9	29
Männer	72	43	19.5	42	186	24	10.4	21	151	38	16.8	30
Körpergröße												
Frauen	92	165	5.8	165	251	168	6.9	168	181	168	6.1	168
Männer	72	180	7.6	180	186	181	7.5	180	151	181	6.1	181
Körpergewicht												
Frauen	92	69	15.3	66.5	251	62	11.2	60	181	63	12.5	59
Männer	72	85	19.3	83	186	77	12.6	75	151	80	11.5	80

Anmerkungen.

*n* = Anzahl; *M* = Mittelwert; *SD* = Standardabweichung; *Md* = Median.

<sup>1</sup> Sechsstufige Antwortskala.

#### 4.3.2 Korrelations-, Reliabilitäts- und Itemanalyse

Bei der Korrelationsanalyse ist unter beiden Bedingungen (6PS und 2PS) und bei beiden Geschlechtern ein ähnliches Muster erkennbar. So zeigt sich ein positiver Zusammenhang (alle angeführten Zusammenhänge signifikant bei  $p < .01$ , falls nicht anders angegeben) zwischen den realen Werten der Merkmale Körpergewicht und Körpergröße (Frauen 2PS:  $r_{pb} = .37$ ; Frauen 6PS:  $r_s = .37$ ; Männer 2PS:  $r_{pb} = .50$ ; Männer 6PS:  $r_s = .48$ ) sowie Körpergewicht und Alter (Frauen 2PS:  $r_{pb} = .24$ ; Frauen 6PS:  $r_s = .29$ ; Männer 2PS:  $r_{pb} = .23$ ; Männer 6PS:  $r_s = .29$ ) und ein schwacher negativer Zusammenhang zwischen Alter und Körpergröße (Frauen 2PS:  $r_{pb} = -.08$  (nicht signifikant bei  $p < .01$ ); Frauen 6PS:  $r_s = -.11$ ; Männer 2PS:  $r_{pb} = -.15$ ; Männer 6PS:  $r_s = -.14$ ). Insgesamt ergeben sich die höchsten positiven Zusammenhänge zwischen den realen Werten eines Merkmals und den Items, die dieses Merkmal erfassen sollen. Es zeigen sich geringe positive Zusammenhänge zwischen manchen Altersitems und den realen Werten des Merkmals Körpergewicht. Es zeigen sich geringe bis mittlere positive Zusammenhänge zwischen Körpergrößenitems und den realen Werten des Merkmals Körpergewicht, umgekehrt ist dieses Muster schwächer ausgeprägt (keine oder nur geringe Zusammenhänge zwischen den Körpergewichtsitems und den

realen Werten des Merkmals Körpergröße). Analog zu den Zusammenhängen zwischen den realen Werten der Merkmale zeigen sich Zusammenhänge zwischen den Items, die diese Merkmale erfassen sollen (geringe überwiegend negative Zusammenhänge zwischen Körpergrößen- und Altersitems; geringe bis mittlere, überwiegend positive Zusammenhänge zwischen Körpergrößen- und Körpergewichtsitems; geringe bis mittlere positive Zusammenhänge zwischen Körpergewichts- und Altersitems).

Auffallend sind Item 6 („Ich mache mir viele Gedanken über das Älterwerden.“) und Item 35 („Wenn ich die Wahl habe zwischen Aufzug und Treppe, dann nehme ich den Aufzug.“). Bei Item 6 zeigt sich kein signifikanter Zusammenhang zwischen gewählten Antwortkategorien und realen Werten (Frauen 2PS:  $r_{pb} = -.02$ ; Frauen 6PS:  $r_s = .09$ ; Männer 2PS:  $r_{pb} = .01$ ; Männer 6PS:  $r_s = .07$ ). Bei Item 35 ist kein oder nur ein geringer Zusammenhang zwischen gewählten Antwortkategorien und realen Werten zu beobachten (Frauen 2PS:  $r_{pb} = .04$ ; (nicht signifikant bei  $p < .01$ ); Frauen 6PS:  $r_s = .17$ ; Männer 2PS:  $r_{pb} = .16$ ; Männer 6PS:  $r_s = .08$ ; (nicht signifikant bei  $p < .01$ )). Es zeigen sich keine oder nur geringe Zusammenhänge zwischen den beiden Items und allen anderen Items. Die Korrelationstabellen finden sich in Anhang C (Tabellen C.7., C.8., C.9. und C.10.).

Tabelle 4.3.5. zeigt die Reliabilitätsschätzungen beider Bedingungen. Dabei ist ersichtlich, dass die Reliabilitäten unter Bedingung 6PS rein deskriptiv höher ausfallen als unter Bedingung 2PS. Unter beiden Bedingungen sind beim Merkmal Körpergröße die Reliabilitäten am höchsten. Unter Bedingung 6PS ergeben sich für das Merkmal Alter die geringsten Reliabilitäten. Unter Bedingung 2PS fallen die Reliabilitäten für die Merkmale Alter und Körpergewicht nahezu identisch aus. Die Reliabilitätsmaßzahlen Kuder-Richardson-20 und McDonalds Omega bzw. Cronbachs Alpha und McDonalds Omega divergieren kaum.

Tabelle 4.3.5.

*Reliabilitäten für die Bedingungen 2PS<sup>1</sup> und 6PS<sup>2</sup>*

Merkmal	2PS <sup>1</sup>		6PS <sup>2</sup>	
	KR-20	$\omega$	$\alpha$	$\omega$
Alter	.79	.79	.85	.85
Körpergröße	.87	.88	.92	.92
Körpergewicht	.78	.79	.88	.89

*Anmerkungen.*KR-20 = Kuder-Richardson-20;  $\alpha$  = Cronbachs Alpha;  $\omega$  = McDonalds Omega.<sup>1</sup> Dichotome Antwortskala.<sup>2</sup> Sechsstufige Antwortskala.

Tabelle 4.3.6. zeigt für beide Bedingungen die Schwierigkeitsindizes, Standardabweichungen und Trennschärfen aller Items. Unter beiden Bedingungen zeichnet sich ein vergleichbares Bild ab. Die Schwierigkeiten und Trennschärfen divergieren deutlich, sowohl innerhalb aller Items als auch innerhalb der Items einer Skala. Die Schwierigkeiten liegen zwischen  $M = .03$  und  $M = .75$  (Item 25 und 22; 2PS) bzw. zwischen  $M = 1.33$  und  $M = 4.41$  (Item 25 und 22; 6PS). Die Trennschärfen liegen zwischen  $r_{i(t-i)} = .12$  und  $r_{i(t-i)} = .66$  (Item 6 und 29; 2PS) bzw. zwischen  $r_{i(t-i)} = .22$  und  $r_{i(t-i)} = .81$  (ebenfalls Item 6 und 29; 6PS).

Die Verteilungen aller Items wurden mithilfe von Histogrammen auf auffällige Merkmale hin untersucht. Bei allen 36 Items wurde das gesamte Spektrum an Antwortkategorien ausgenutzt. Beim Merkmal Alter zeigen sieben Items (1, 2, 4, 5, 7, 8 und 12) eine linkssteile Verteilung, zwei Items (6 und 11) eine Gleichverteilung, ein Item (3) eine symmetrische Verteilung und ein Item (9) eine leicht bimodale Verteilung. Beim Merkmal Körpergröße weisen acht Items (13, 14, 15, 17, 18, 19, 20 und 24) eine linkssteile Verteilung auf, zwei Items (22 und 23) eine rechtssteile Verteilung und zwei Items (16 und 21) eine leicht bimodale Verteilung. Beim Merkmal Körpergewicht ist bei neun Items (25, 26, 27, 28, 29, 30, 34, 35 und 36) eine linkssteile Verteilung ersichtlich, bei zwei Items (31 und 32) eine symmetrische Verteilung und bei einem Item (33) eine Gleichverteilung. Die Histogramme finden sich in Anhang C (Abbildungen C.14., C.15., C.16., C.17., C.18. und C.19.).

Tabelle 4.3.6.

Mittelwerte ( $M$ ), Standardabweichungen ( $SD$ ) und Trennschärfen ( $r_{i(t-i)}$ ) für die Bedingungen 2PS<sup>1</sup> und 6PS<sup>2</sup>

Merkmal und Item-Nummer		2PS <sup>1</sup>			6PS <sup>2</sup>		
		$M$	$SD$	$r_{i(t-i)}$	$M$	$SD$	$r_{i(t-i)}$
Alter							
1	(D, P)	.17	.37	.29	2.10	1.44	.43
2	(D)	.23	.42	.44	2.40	1.43	.60
3	(D, P)	.55	.50	.38	3.84	1.25	.43
4	(D, P)	.23	.42	.54	2.36	1.34	.61
5	(D, P)	.33	.47	.55	2.66	1.48	.63
6	(D, P)	.42	.49	.12	3.39	1.54	.22
7	(D)	.21	.41	.61	2.36	1.58	.70
8	(D, P)	.22	.42	.42	2.30	1.42	.52
9	(D)	.41	.49	.49	2.92	1.96	.54
10	(D, P)*	.34	.48	.55	2.98	1.49	.56
11	(D, P)	.43	.50	.33	3.15	1.72	.45
12	(P)	.20	.40	.49	2.24	1.42	.64
Körpergröße							
13	(D, P)	.10	.29	.42	1.77	1.33	.48
14	(D)	.34	.47	.63	2.76	1.68	.76
15	(D)	.13	.33	.49	2.01	1.36	.62
16	(D, P)	.65	.48	.61	4.04	1.92	.73
17	(D, P)	.39	.49	.60	2.81	1.89	.65
18	(D)	.17	.38	.58	1.98	1.53	.65
19	(D, P)	.34	.47	.64	2.82	1.80	.73
20	(D, P)	.19	.39	.42	2.10	1.57	.58
21	(D, P)*	.65	.48	.61	4.13	1.89	.70
22	(D, P)*	.75	.43	.52	4.41	1.59	.60
23	(D, P)*	.70	.46	.59	4.24	1.77	.69
24	(P)	.25	.43	.65	2.39	1.60	.72
Körpergewicht							
25	(D)	.03	.18	.23	1.33	0.89	.39
26	(D)	.23	.42	.32	2.32	1.53	.46
27	(D, P)	.16	.37	.58	2.33	1.42	.73
28	(D)	.11	.31	.57	1.97	1.43	.80
29	(D)	.19	.40	.66	2.38	1.53	.81
30	(D, P)	.31	.46	.62	2.65	1.77	.76
31	(D, P)*	.46	.50	.49	3.26	1.62	.64
32	(D)*	.31	.46	.44	2.98	1.52	.55
33	(D)*	.49	.50	.25	3.45	1.59	.35



Tabelle 4.3.6. (Fortsetzung)

Mittelwerte ( $M$ ), Standardabweichungen ( $SD$ ) und Trennschärfen ( $r_{i(t-i)}$ ) für die Bedingungen 2PS<sup>1</sup> und 6PS<sup>2</sup>

Merkmal und Item-Nummer	2PS <sup>1</sup>			6PS <sup>2</sup>		
	$M$	$SD$	$r_{i(t-i)}$	$M$	$SD$	$r_{i(t-i)}$
34 (D)	.07	.26	.32	1.95	1.25	.58
35 (P)	.31	.46	.17	2.91	1.71	.24
36 (P)	.37	.48	.46	2.81	1.73	.62

Anmerkungen.

$M$  = Mittelwert;  $SD$  = Standardabweichung;  $r_{i(t-i)}$  = Trennschärfe.

<sup>1</sup> Dichotome Antwortskala.

<sup>2</sup> Sechsstufige Antwortskala.

D = Items hervorgegangen aus dem deduktiven Testkonstruktionsansatz.

P = Items hervorgegangen aus dem Prototypenansatz.

\* Items mit negativer Polung; die Nummerierung der Items entspricht der Nummerierung in Tabelle C.6.

### 4.3.3 Explorative und Konfirmatorische Faktorenanalyse

Die von der Parallelanalyse ermittelte optimale Faktorenzahl lag bei sechs (2PS) bzw. fünf Faktoren (6PS). Der Versuch, die sechs- bzw. fünffaktoriellen Lösungen zu interpretieren, scheiterte. Die entsprechenden Ergebnisse finden sich in Anhang C (Tabellen C.11. und C.12.).

Die Tabellen 4.3.7. und 4.3.8. zeigen die Ergebnisse der EFA mit drei Faktoren für beide Bedingungen. Bei den Merkmalen Alter und Körpergröße ergeben sich unter Bedingung 6PS bei drei Items geringe Nebenladungen, unter Bedingung 2PS bei zwölf Items. Beim Merkmal Körpergewicht zeigen sich unter beiden Bedingungen die höchsten Nebenladungen. Item 33 („Zwischen den Armlehnen in Flugzeugsitzen habe ich viel Platz.“) lädt unter beiden Bedingungen in nahezu identischer Höhe auf zwei Faktoren (Faktor 1 und Faktor 2). Bei Item 6 und Item 35 ergeben sich insgesamt die geringsten Ladungen und Kommunalitäten nahe Null (Item 6:  $h^2 = .019$  (2PS) bzw.  $h^2 = .071$  (6PS); Item 35:  $h^2 = .050$  (2PS) bzw.  $h^2 = .066$  (6PS)).

Tabelle 4.3.7.

*Explorative Faktorenanalyse mit drei Faktoren (Maximum Likelihood, Promax) – Bedingungen 2PS<sup>1</sup> und 6PS<sup>2</sup>*

Merkmal und Item- Nummer		2PS <sup>1</sup>				6PS <sup>2</sup>			
		Faktor 1	Faktor 2	Faktor 3	$h^2$	Faktor 1	Faktor 2	Faktor 3	$h^2$
Alter									
1	(D, P)		.23	<b>.35</b>	.221		.16	<b>.47</b>	.318
2	(D)			<b>.52</b>	.272			<b>.75</b>	.566
3	(D, P)			<b>.68</b>	.441			<b>.45</b>	.178
4	(D, P)			<b>.68</b>	.453			<b>.74</b>	.567
5	(D, P)			<b>.67</b>	.416			<b>.78</b>	.592
6	(D, P)			<b>.11</b>	.019			<b>.25</b>	.071
7	(D)		-.16	<b>.99</b>	.907			<b>.77</b>	.559
8	(D, P)	-.17		<b>.57</b>	.384			<b>.62</b>	.368
9	(D)			<b>.77</b>	.589			<b>.57</b>	.310
10	(D, P)*			<b>.85</b>	.697			<b>.61</b>	.358
11	(D, P)			<b>.44</b>	.191			<b>.51</b>	.268
12	(P)			<b>.65</b>	.436			<b>.75</b>	.606
Körpergröße									
13	(D, P)	<b>.72</b>	.17	-.17	.629	<b>.60</b>			.384
14	(D)	<b>.85</b>		-.10	.716	<b>.86</b>			.721
15	(D)	<b>.79</b>			.620	<b>.74</b>			.555
16	(D, P)	<b>.82</b>		.13	.644	<b>.84</b>	-.13		.675
17	(D, P)	<b>.82</b>		.13	.668	<b>.76</b>			.583
18	(D)	<b>.85</b>		-.15	.762	<b>.80</b>			.654
19	(D, P)	<b>.85</b>			.747	<b>.82</b>			.682
20	(D, P)	<b>.68</b>	-.15		.440	<b>.71</b>			.502
21	(D, P)*	<b>.81</b>			.656	<b>.81</b>		-.17	.663
22	(D, P)*	<b>.76</b>		.23	.583	<b>.67</b>			.437
23	(D, P)*	<b>.80</b>		.15	.656	<b>.77</b>			.591
24	(P)	<b>.87</b>		-.18	.805	<b>.83</b>			.667
Körpergewicht									
25	(D)	.20	<b>.51</b>		.353	.14	<b>.46</b>	.18	.361
26	(D)	-.18	<b>.68</b>	-.31	.382		<b>.59</b>	-.11	.294
27	(D, P)		<b>.79</b>		.656		<b>.81</b>		.667
28	(D)	-.11	<b>.91</b>		.858	-.18	<b>.98</b>		.908
29	(D)		<b>.93</b>		.910		<b>.92</b>		.826
30	(D, P)	-.22	<b>.91</b>		.793	-.23	<b>.91</b>		.801
31	(D, P)*		<b>.68</b>		.499		<b>.75</b>		.523
32	(D)*	.16	<b>.53</b>	.10	.398		<b>.57</b>		.359

Tabelle 4.3.7. (Fortsetzung)

Explorative Faktorenanalyse mit drei Faktoren (Maximum Likelihood, Promax) – Bedingungen 2PS<sup>1</sup> und 6PS<sup>2</sup>

Merkmal und Item- Nummer	2PS <sup>1</sup>				6PS <sup>2</sup>			
	Faktor 1	Faktor 2	Faktor 3	$h^2$	Faktor 1	Faktor 2	Faktor 3	$h^2$
33 (D)*	<b>.31</b>	.30		.243	.29	<b>.30</b>		.222
34 (D)	.27	<b>.51</b>		.405	.31	<b>.60</b>		.541
35 (P)		<b>.22</b>		.050		<b>.25</b>		.066
36 (P)	-.25	<b>.74</b>		.508	-.21	<b>.79</b>		.567

Anmerkungen.

 $h^2$  = Kommunalität des Items; Ladungen mit Betrag < .10 wurden unterdrückt.

Die stärkste Ladung für jedes Item ist fett gedruckt.

<sup>1</sup> Dichotome Antwortskala.<sup>2</sup> Sechsstufige Antwortskala.

D = Items hervorgegangen aus dem deduktiven Testkonstruktionsansatz.

P = Items hervorgegangen aus dem Prototypenansatz.

\* Items mit negativer Polung; die Nummerierung der Items entspricht der Nummerierung in Tabelle C.6.

Tabelle 4.3.8.

Faktorkorrelationen EFA mit drei Faktoren (Maximum Likelihood, Promax) – Bedingungen 2PS<sup>1</sup> und 6PS<sup>2</sup>

Faktor	2PS <sup>1</sup>			6PS <sup>2</sup>		
	1	2	3	1	2	3
1	-			-		
2	.28	-		.22	-	
3	-.02	.32	-	.03	.41	-

Anmerkungen.

EFA = Explorative Faktorenanalyse.

<sup>1</sup> Dichotome Antwortskala.<sup>2</sup> Sechsstufige Antwortskala.

Bei der CFA werden die Daten unter Bedingung 2PS von Modell 1 unbefriedigend beschrieben ( $\chi^2(591, n = 921) = 2\,405.19, p < .001$ ; CFI = .90; RMSEA = .06 (90 % KI: 0.055, 0.060); SRMR = .06). Die standardisierten Ladungen liegen zwischen  $\lambda = .128$  (Item 6) und  $\lambda = .755$  (Item 29). Auffallend niedrige Ladungen zeigen sich bei Item 6 ( $\lambda = .128$ ), Item 25 ( $\lambda = .264$ ), Item 26 ( $\lambda = .280$ ) und Item 35 ( $\lambda = .181$ ). Es ergeben sich signifikante Korrelationen zwischen den latenten Faktoren Alter und Körpergewicht ( $r = .28, z = 6.80, p < .001$ ) sowie Körpergröße und Körpergewicht ( $r = .23, z = 5.75, p < .001$ ). Der

Zusammenhang zwischen Alter und Körpergröße ist nicht signifikant von Null verschieden ( $r = -.06$ ,  $z = -1.66$ ,  $p = .10$ ).

Unter Bedingung 6PS zeigt sich ein ähnliches Bild wie unter Bedingung 2PS ( $\chi^2(591, n = 933) = 3\,104.28$ ,  $p < .001$ ; CFI = .92; RMSEA = .07 (90 % KI: 0.065, 0.070); SRMR = .07). Die standardisierten Ladungen liegen zwischen  $\lambda = .252$  (Item 6) und  $\lambda = .880$  (Item 29). Auffallend niedrige Ladungen zeigen sich bei Item 6 ( $\lambda = .252$ ) und Item 35 ( $\lambda = .253$ ). Es ergeben sich signifikante Korrelationen zwischen den latenten Faktoren Alter und Körpergewicht ( $r = .39$ ,  $z = 10.67$ ,  $p < .001$ ) sowie Körpergröße und Körpergewicht ( $r = .15$ ,  $z = 4.26$ ,  $p < .001$ ). Der Zusammenhang zwischen Alter und Körpergröße ist nicht signifikant von Null verschieden ( $r = -.004$ ,  $z = -0.11$ ,  $p = .91$ ).

Unter beiden Bedingungen treten bei den Schätzungen für Modell 2 Heywood-Cases in Form negativer Fehlervarianzen auf (Heywood, 1931). Wie in Modell 1 ersichtlich ist, korrelieren nicht alle drei latenten Faktoren (Alter, Körpergröße und Körpergewicht) miteinander, weshalb ein übergeordneter Generalfaktor, der diese Korrelationen erklärt, nicht angenommen werden kann. Auf eine Ergebnisdarstellung von Modell 2 wird verzichtet. Die Ergebnisse von Modell 1 finden sich in Anhang C (Tabelle C.13. sowie Abbildungen C.20. und C.21.).

#### 4.3.4 Probabilistische Testmodelle

Bei den probabilistischen Testmodellen beschreibt unter Bedingung 2PS bei allen Merkmalen das 2PL-Modell im Vergleich zum 1PL-Modell die Daten besser (Alter:  $\Delta AIC = 293.4$ ;  $\Delta BIC = 240.3$ ;  $\chi^2(24) = 315.39$ ,  $p < .001$ ; Körpergröße:  $\Delta AIC = 70.7$ ;  $\Delta BIC = 17.6$ ;  $\chi^2(24) = 92.67$ ,  $p < .001$ ; Körpergewicht:  $\Delta AIC = 436.4$ ;  $\Delta BIC = 383.3$ ;  $\chi^2(24) = 458.41$ ,  $p < .001$ ). Zur Beurteilung der globalen Modellgültigkeit ergibt unter Bedingung 2PS für das Merkmal Alter der Pearson Chi-Quadrat-Test beim 1PL-Modell bei approximativer Chi-Quadrat-Verteilung  $PE_{(df = 4\,082)} = 6\,217.31$ ,  $p < .001$ , die Bootstrap-Überschreitungswahrscheinlichkeit beträgt  $p = .003$ . Beim 2PL-Modell ergibt der Pearson Chi-Quadrat-Test bei approximativer Chi-Quadrat-Verteilung  $PE_{(df = 4\,071)} = 5\,361.35$ ,  $p < .001$ , die Bootstrap-Überschreitungswahrscheinlichkeit beträgt  $p = .007$ . Für das Merkmal Körpergröße ergibt beim 1PL-Modell der Pearson Chi-Quadrat-Test bei approximativer Chi-Quadrat-Verteilung  $PE_{(df = 4\,082)} = 88\,031.006$ ,  $p < .001$ , die Bootstrap-Überschreitungswahrscheinlichkeit beträgt  $p = .003$ . Beim 2PL-Modell ergibt der Pearson Chi-Quadrat-Test bei approximativer

Chi-Quadrat-Verteilung  $PE_{(df = 4\ 071)} = 208\ 602\ 855$ ,  $p < .001$ , die Bootstrap-Überschreitungswahrscheinlichkeit beträgt  $p = .003$ . Für das Merkmal Körpergewicht ergibt beim 1PL-Modell der Pearson Chi-Quadrat-Test bei approximativer Chi-Quadrat-Verteilung  $PE_{(df = 4\ 082)} = 8\ 090.48$ ,  $p < .001$ , die Bootstrap-Überschreitungswahrscheinlichkeit beträgt  $p = .02$ . Beim 2PL-Modell ergibt der Pearson Chi-Quadrat-Test bei approximativer Chi-Quadrat-Verteilung  $PE_{(df = 4\ 071)} = 10\ 148.97$ ,  $p < .001$ , die Bootstrap-Überschreitungswahrscheinlichkeit beträgt  $p = .02$ . Auf Basis der berichteten Ergebnisse muss unter Bedingung 2PS die globale Modellgültigkeit für alle Modelle verworfen werden.

Unter Bedingung 6PS beschreibt bei allen Merkmalen das GPCM im Vergleich zum PCM die Daten besser (Alter:  $\Delta AIC = 452.3$ ;  $\Delta BIC = 399.0$ ;  $\chi^2(72) = 474.26$ ,  $p < .001$ ; Körpergröße:  $\Delta AIC = 129.8$ ;  $\Delta BIC = 76.6$ ;  $\chi^2(72) = 151.82$ ,  $p < .001$ ; Körpergewicht:  $\Delta AIC = 1\ 203.2$ ;  $\Delta BIC = 1\ 150.0$ ;  $\chi^2(72) = 1\ 225.20$ ,  $p < .001$ ). Beim PCM ergeben sich nur bei vier Items vollständig geordnete Schwellenparameter (Alter: Item 3; Körpergewicht: Items 27, 32 und 33), beim GPCM ist dies bei sieben Items der Fall (Alter: Item 3; Körpergröße: Item 14; Körpergewicht: Items 27, 28, 29, 33 und 34). Die Abbildungen 4.3.1. und 4.3.2. zeigen beispielhaft Item Characteristic Curves für Items mit geordneten bzw. ungeordneten Schwellenparametern (GPCM).

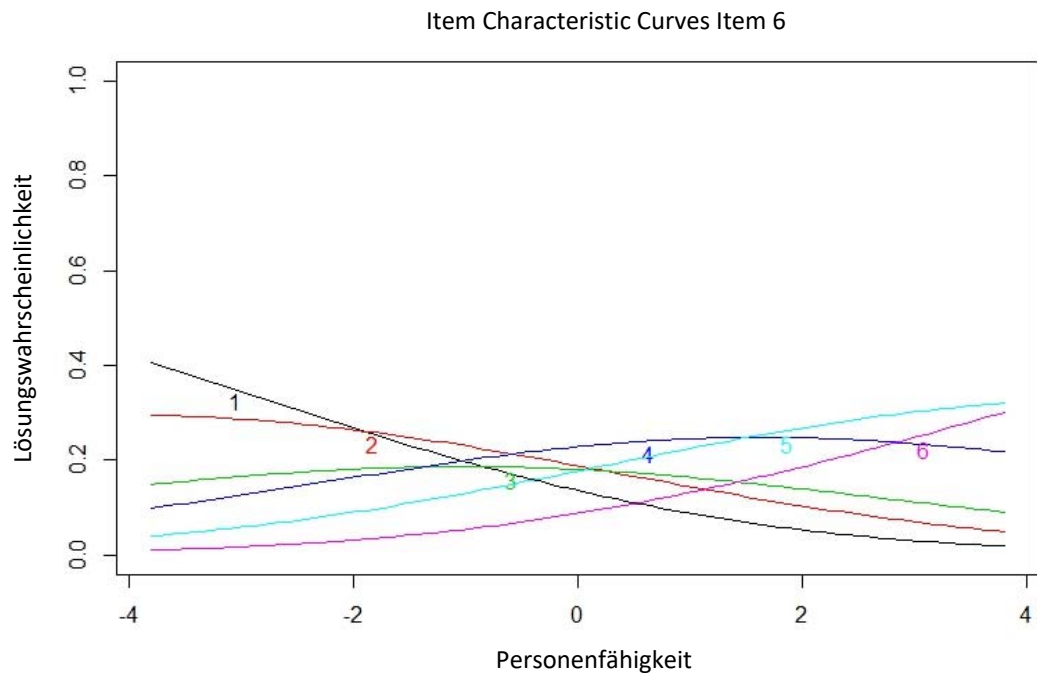


Abbildung 4.3.1. Item Characteristic Curves für Item 6 („Ich mache mir viele Gedanken über das Älterwerden.“) als Beispiel für ungeordnete Schwellen. Die schwarze Kurve entspricht der Kategorienwahrscheinlichkeit der Kategorie 1, rot Kategorie 2, grün Kategorie 3, blau Kategorie 4, türkis Kategorie 5 und magenta Kategorie 6; GPCM.

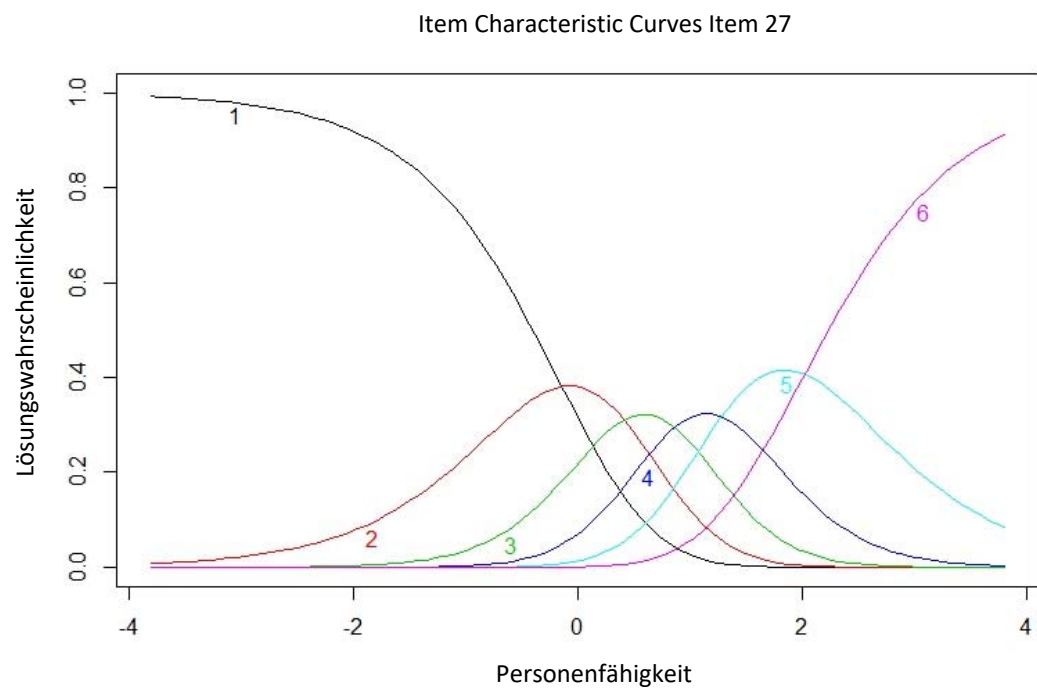


Abbildung 4.3.2. Item Characteristic Curves für Item 27 („Bei Hosen habe ich eine große Bundweite.“) als Beispiel für geordnete Schwellen. Die schwarze Kurve entspricht der Kategorienwahrscheinlichkeit der Kategorie 1, rot Kategorie 2, grün Kategorie 3, blau Kategorie 4, türkis Kategorie 5 und magenta Kategorie 6; GPCM.

Zur Beurteilung der globalen Modellgültigkeit ergibt unter Bedingung 6PS für das Merkmal Alter der Pearson Chi-Quadrat-Test beim PCM bei approximativer Chi-Quadrat-Verteilung  $PE_{(df = 2\ 176\ 782\ 274)} = 688\ 573\ 558\ 164$ ,  $p < .001$ , die Bootstrap-Überschreitungswahrscheinlichkeit beträgt  $p = .003$ . Beim GPCM ergibt der Pearson Chi-Quadrat-Test bei approximativer Chi-Quadrat-Verteilung  $PE_{(df = 2\ 176\ 782\ 263)} = 209\ 936\ 389\ 945$ ,  $p < .001$ , die Bootstrap-Überschreitungswahrscheinlichkeit beträgt  $p = .003$ . Für das Merkmal Körpergröße ergibt beim PCM der Pearson Chi-Quadrat-Test bei approximativer Chi-Quadrat-Verteilung  $PE_{(df = 2\ 176\ 782\ 274)} = 3.496468e^{+16}$ ,  $p < .001$ , die Bootstrap-Überschreitungswahrscheinlichkeit beträgt  $p = .003$ . Beim GPCM ergibt der Pearson Chi-Quadrat-Test bei approximativer Chi-Quadrat-Verteilung  $PE_{(df = 2\ 176\ 782\ 263)} = 2.226835e^{+16}$ ,  $p < .001$ , die Bootstrap-Überschreitungswahrscheinlichkeit beträgt  $p = .003$ . Für das Merkmal Körpergewicht ergibt beim PCM der Pearson Chi-Quadrat-Test bei approximativer Chi-Quadrat-Verteilung  $PE_{(df = 2\ 176\ 782\ 274)} = 333\ 677\ 500\ 290$ ,  $p < .001$ , die Bootstrap-Überschreitungswahrscheinlichkeit beträgt  $p = .003$ . Beim GPCM ergibt der Pearson Chi-Quadrat-Test bei approximativer Chi-Quadrat-Verteilung  $PE_{(df = 2\ 176\ 782\ 263)} = 283\ 262\ 128\ 373$ ,  $p < .001$ , die Bootstrap-Überschreitungswahrscheinlichkeit beträgt  $p = .003$ . Auf Basis der berichteten Ergebnisse muss unter Bedingung 6PS die globale Modellgültigkeit für alle Modelle ebenfalls verworfen werden. Die Ergebnisse der probabilistischen Testmodelle finden sich in Anhang C (Tabellen C.14., C.15. und C.16.).

#### 4.3.5 Regressionsanalyse

Tabelle 4.3.9. zeigt die Ergebnisse der einfachen linearen Regressionen. Dabei ist ersichtlich, dass unter Bedingung 2PS die Determinationskoeffizienten  $R^2_{SW}$ ,  $R^2_{FW}$  und  $R^2_{PP}$  innerhalb der Merkmale Körpergröße und Körpergewicht annähernd gleich ausfallen. Zwischen den Merkmalen ergeben sich bei Körpergröße die höchsten und bei Körpergewicht (in Kilogramm) die geringsten Determinationskoeffizienten. Beim Merkmal Alter klären Summenwerte einen geringeren Varianzanteil auf als Faktorwerte und Personenparameter.

Unter Bedingung 6PS klären Personenparameter bei allen Merkmalen weniger Varianz auf als Summenwerte und Faktorwerte.  $R^2_{SW}$  und  $R^2_{FW}$  fallen innerhalb aller Merkmale nahezu identisch aus. Zwischen den Merkmalen ergeben sich insgesamt bei Körpergröße die

höchsten und bei Körpergewicht (in Kilogramm) die niedrigsten Determinationskoeffizienten. Ein Vergleich der Bedingungen 2PS und 6PS zeigt keine eindeutigen Vorteile für ein Antwortskalenformat. So fallen die Determinationskoeffizienten  $R^2_{SW}$ ,  $R^2_{FW}$  und  $R^2_{PP}$  bei den Merkmalen Körpergröße und Körpergewicht (in Kilogramm) annähernd gleich aus. Beim Merkmal Alter schneidet die dichotome Antwortskala im Vergleich zur sechsstufigen Antwortskala leicht besser ab, beim Merkmal Körpergewicht (BMI) zeigt sich ein umgekehrtes Bild. Ein Vergleich beider Bedingungen über alle Merkmale konnte keine eindeutige Aussage darüber liefern, ob Summenwerte, Faktorwerte oder Personenparameter die realen Werte in den Regressionsanalysen besser vorhersagen.

Tabelle 4.3.9.

*Ergebnisse einfache lineare Regressionen für Alter, Körpergröße und Körpergewicht – Bedingungen 2PS<sup>1</sup> u. 6PS<sup>2</sup>*

Merkmal	2PS <sup>1</sup>			6PS <sup>2</sup>		
	$R^2_{SW}$	$R^2_{FW}$	$R^2_{PP}$	$R^2_{SW}$	$R^2_{FW}$	$R^2_{PP}$
Alter	.53	.61	.57	.52	.49	.43
Körpergröße	.70	.71	.71	.73	.73	.72
Körpergewicht (Kilogramm)	.44	.44	.41	.46	.46	.42
Körpergewicht (BMI)	.51	.54	.51	.58	.60	.55

*Anmerkungen.*

$R^2$  = Determinationskoeffizient; *SW* = Summenwerte; *FW* = Bartlett Faktorwerte aus der konfirmatorischen Faktorenanalyse (Modell 1); *PP* = Personenparameter (expected a posteriori Schätzer; 2PL-Modell bzw. GPCM). Alter in Jahren; Körpergröße in Zentimetern; Körpergewicht in Kilogramm bzw. BMI.

<sup>1</sup> Dichotome Antwortskala.

<sup>2</sup> Sechsstufige Antwortskala.

#### 4.3.6 Ergebnisse im Kontext der Forschungsfragen

*Forschungsfragen 1 und 2.* Unter Abschnitt 4.1.4 wurde die Frage aufgeworfen, wie die Skalen bezüglich Itemschwierigkeit, Itemtrennschärfe und Reliabilität zu beurteilen sind. Schwierigkeiten und Trennschärfen variieren deutlich innerhalb aller Skalen. Besonders niedrig fallen die Trennschärfen bei den Items 6 und 35 aus. Rein deskriptiv kann bei keiner Skala von einer Passung eines tau-äquivalenten Messmodells ausgegangen werden (Bühner,



2011). Die Reliabilitäten fallen unter Bedingung 6PS höher aus als unter Bedingung 2PS. Unter beiden Bedingungen sind beim Merkmal Körpergröße die Reliabilitäten am höchsten.

*Forschungsfrage 3.* Unter Abschnitt 4.1.4 wurde die Frage aufgeworfen, ob sich auffallende Korrelationen zwischen den Items (gewählte Antwortkategorie) und realen Werten sowie zwischen den Items unterschiedlicher Skalen ergeben. Es zeigen sich geringe bis mittlere positive Zusammenhänge zwischen den Körpergrößenitems und den realen Werten des Merkmals Körpergewicht. Umgekehrt zeigen sich keine oder nur geringe Zusammenhänge zwischen den Körpergewichtsitems und den realen Werten des Merkmals Körpergröße. Außerdem sind geringe bis mittlere, meist positive Zusammenhänge zwischen Körpergrößen- und Körpergewichtsitems zu beobachten sowie geringe bis mittlere positive Zusammenhänge zwischen Körpergewichts- und Altersitems. Auffallend sind Item 6 und 35. Beide Items weisen keine oder nur geringe Zusammenhänge zwischen gewählter Antwortkategorie und den realen Werten auf.

*Forschungsfrage 4.* Unter Abschnitt 4.1.4 wurde die Frage aufgeworfen, ob sich bei einer EFA eine dreifaktorielle Lösung ergibt, bei der alle Items auf den entsprechend vorgesehenen Faktor laden. Die Parallelanalyse ermittelte sechs- (2PS) bzw. fünffaktorielle (6PS), inhaltlich schwer zu interpretierende Lösungen. Bei der dreifaktoriellen Lösung lädt Item 33 in fast gleicher Höhe auf zwei Faktoren. Item 6 und 35 weisen Kommunalitäten nahe Null auf.

*Forschungsfrage 5.* Unter Abschnitt 4.1.4 wurde die Frage aufgeworfen, welches Modell im Rahmen einer CFA die Daten besser abbildet. Bei beiden Antwortskalenformaten kann weder Modell 1 noch Modell 2 die Daten zufriedenstellend beschreiben.

*Forschungsfrage 6.* Unter Abschnitt 4.1.4 wurde die Frage aufgeworfen, durch welches der probabilistischen Testmodelle die Daten besser beschrieben werden. Im relativen Modellvergleich beschreibt bei allen Merkmalen das 2PL-Modell bzw. GPCM die Daten besser als das 1PL-Modell bzw. das PCM. Die globale Modellgültigkeit muss unter allen Bedingungen für alle Modelle verworfen werden.

*Forschungsfrage 7.* Unter Abschnitt 4.1.4 wurde die Frage aufgeworfen, in welcher Höhe die Determinationskoeffizienten  $R^2$  bei einfachen linearen Regressionen ausfallen, mit Summenwerten, Faktorwerten oder Personenparametern als unabhängige Variablen und Alter, Körpergröße oder Körpergewicht bzw. BMI als abhängige Variablen. Bei beiden

Antwortskalenformaten fallen beim Merkmal Körpergröße die Determinationskoeffizienten  $R^2$  am höchsten und beim Merkmal Körpergewicht (in Kilogramm) am geringsten aus. Im Vergleich über beide Bedingungen und alle Merkmale wird nicht eindeutig ersichtlich, ob Summenwerte, Faktorwerte oder Personenparameter mehr Varianz aufklären.

*Forschungsfrage 8.* Unter Abschnitt 4.1.4 wurde die Frage aufgeworfen, ob beim Merkmal Körpergewicht die Regressionsanalysen unterschiedlich ausfallen, wenn sie auf Basis des BMI anstatt des absoluten Körpergewichts durchgeführt werden. Bei beiden Antwortskalenformaten fallen die Determinationskoeffizienten bei Regressionsanalysen auf Basis des BMI höher aus als auf Basis des absoluten Körpergewichts.

*Forschungsfrage 9.* Unter Abschnitt 4.1.4 wurde die Frage aufgeworfen, ob sich das zwei- oder sechsstufige Antwortskalenformat besser eignet, um die realen Werte in den Regressionsanalysen vorherzusagen. Es zeigte sich kein eindeutiges Muster, dass ein Antwortskalenformat über eine bessere Prädiktionsgüte verfügt als das andere.

#### **4.4 Diskussion Studie III**

##### **4.4.1 Zusammenfassung der Ergebnisse**

Zielsetzung von Studie III war die Überprüfung der Frage, wie gut psychologische Skalen in der Lage sind, die physikalisch messbaren Personenmerkmale Alter, Körpergröße und Körpergewicht zu erfassen. Die Skalenkonstruktion folgte den in der Fachliteratur gegebenen Empfehlungen zur Fragebogenkonstruktion. Ein dichotomes Antwortformat wurde einem sechsstufigen Antwortformat gegenübergestellt.

Die Reliabilitätsschätzungen fielen unter Bedingung 6PS etwas höher aus als unter Bedingung 2PS. Unter beiden Bedingungen waren beim Merkmal Körpergröße die Reliabilitäten am höchsten. Zwischen den Körpergrößentitems und den realen Werten des Merkmals Körpergewicht zeigten sich auffällige positive Zusammenhänge, umgekehrt war dieses Muster deutlich weniger ausgeprägt. Unter keiner Bedingung und bei keinem Merkmal konnten die spezifizierten Modelle (EFA, CFA, probabilistische Testmodelle) die Daten angemessen beschreiben. Die Items 6 und 35 wiesen die schlechtesten psychometrischen Eigenschaften auf. Unter beiden Bedingungen konnte beim Merkmal Körpergröße die meiste Varianz aufgeklärt werden, beim Merkmal Körpergewicht (in Kilogramm) die geringste. Ein Vergleich beider Bedingungen über alle Merkmale ergab keine eindeutige Antwort auf die Frage, ob Summenwerte, Faktorwerte oder Personenparameter

mehr Varianz aufklären. Regressionsanalysen auf Basis des BMI ergaben einen höheren Anteil aufgeklärter Varianz als Regressionsanalysen auf Basis des absoluten Körpergewichts. Der Vergleich beider Antwortskalenformate konnte keine eindeutige Aussage darüber liefern, welches der Skalenformate die realen Werte in den Regressionsanalysen besser vorhersagen kann.

Die Interpretation der Ergebnisse bezieht sich vornehmlich auf die Hauptbefunde.

#### **4.4.2 Interpretation und Implikationen der Ergebnisse**

Die in dieser Arbeit konstruierten Skalen konnten Varianz zwischen  $R^2 = .41$  und  $R^2 = .73$  aufklären. Psychologische Effekte sind im Allgemeinen sehr gering, mit einer geschätzten mittleren Korrelation von  $|r| = .16$  (Bosco, Aguinis, Singh, Field & Pierce, 2015). Damit zeigten sich starke, in der Psychologie selten zu beobachtende, Effekte (Bühner & Ziegler, 2017; Cohen, 1988; Cohen, 1992; Ellis, 2010). Wie in der vergleichbaren von Bortolotti und Kollegen (2013) durchgeführten Studie, sind Messungen physikalischer Merkmale mit psychologischen Skalen grundsätzlich möglich und liefern für psychologische Maßstäbe gute Ergebnisse.

Antwortkategorien von Fragebogenitems im Likert-Skalenformat werden von Menschen unterschiedlich verwendet (van Vaerenbergh & Thomas, 2013). Manche Personen neigen dazu, mittlere Kategorien anzukreuzen, während andere eher extremere Kategorien bevorzugen. Dies geschieht unabhängig von der zu messenden latenten Eigenschaft und wird als *Tendenz zur Mitte* bzw. *extremer Antwortstil* bezeichnet. Bei diesen Antwortstilen handelt es sich um zeitlich stabile Persönlichkeitseigenschaften (Weijters et al., 2010; Wetzel, Lüdtke, Zettler & Böhnke, 2016), die in Fragebögen zu unterschiedlichen Konstrukten gleichermaßen auftreten (Wetzel, Carstensen & Böhnke, 2013). Eine Nichtberücksichtigung dieser Antwortstile kann dazu führen, dass Korrelationen zwischen psychologischen Konstrukten (Baumgartner & Steenkamp, 2001) sowie Varianzen von Skalenwerten (Wetzel & Carstensen, 2017) überschätzt werden. Als wirkungsvolles Mittel zur Vermeidung extremer Antwortstile werden Items mit dichotomem Antwortformat diskutiert (Bühner, 2011). In einer ersten empirischen Studie hat sich dieses Vorgehen als wirksame Strategie bestätigt (Pargent, 2017). Viele Anwender stehen dichotomen Antwortformaten bisher skeptisch gegenüber. Sie befürchten vor allem einen damit einhergehenden Informationsverlust, der sich aber in Studie III nicht zeigte. Beim Vergleich

der Ergebnisse beider Antwortskalenformate fielen die Reliabilitäten unter Bedingung 6PS höher aus als unter Bedingung 2PS. Die Unterschiede sind jedoch teilweise so gering, dass sich daraus nicht eindeutig schlussfolgern lässt, dass eines der beiden Skalenformate die drei untersuchten Merkmale präziser erfasst. Außerdem zeigte sich kein eindeutiges Muster dafür, dass ein Antwortskalenformat die realen Werte in den Regressionsanalysen besser vorhersagen kann. Bereits Studie II ergab keinen klaren Hinweis auf den besseren Erhalt der Relationen realer Werte bei einer der untersuchten Skalenbreiten. Die vorliegenden Ergebnisse reihen sich somit in diese Vorbefunde ein. Damit werden Thesen gestützt, die besagen, dass breitere Antwortskalen lediglich eine Scheingenauigkeit suggerieren (Porst, 2014) und dass dichotome Antwortformate nicht zwangsläufig mit einem Informationsverlust einhergehen (Finn et al., 2015). Für eine präzise Messung psychologischer Konstrukte müssten extreme Antwortstile vermieden werden. Im Rahmen der psychologischen Testkonstruktion sollten deshalb dichotome Antwortformate statt polytomer Antwortformate verwendet werden. Der oft befürchtete Informationsverlust bei dichotomen Items spielt vermutlich eine untergeordnete Rolle.

Bei den drei untersuchten Personenmerkmalen ergab sich beim Merkmal Körpergröße der höchste Anteil aufgeklärter Varianz. Somit konnte das Merkmal Körpergröße am besten auf psychologischen Skalen abgebildet werden. Bereits im Pilottest wiesen Körpergrößensitems bessere psychometrische Eigenschaften auf als Alters- und Körpergewichtsitems. Die Items aller drei Skalen wurden nach demselben Vorgehen und mit demselben Aufwand konstruiert. Deshalb können Unterschiede bei Konstruktion und Formulierung, die zu inhaltlich und sprachlich ausgereifteren Items hätten führen können, ausgeschlossen werden. Die Determinationskoeffizienten  $R^2$  fallen ähnlich aus wie bei Bortolotti und Kollegen ( $R^2 = .737$ ; 2013, S. 2 356), obwohl die dort verwendeten Items andere inhaltliche Aspekte und Formulierungen aufwiesen als die in Studie III. Bereits die Ergebnisse in Eichhorn (2016) und in Studie II legten nahe, dass bei der Selbstbewertung der Körpergröße das subjektive und objektive Körperbild geringer divergieren als bei anderen körperbezogenen Merkmalen. So zeigte ein innerer Perspektivenwechsel, induziert durch unterschiedliche Itemformulierung, keinen Effekt. Dass sich Körpergröße im Vergleich zu Alter und Körpergewicht präziser auf einer psychologischen Skala erfassen lässt, ist damit konsistent zu den Vorbefunden. Wie bereits in Studie II unter Abschnitt 3.4.2 diskutiert,

unterliegt die Selbstbewertung der Merkmale Alter und Körpergewicht individuellen Einflussfaktoren, wie beispielsweise dem subjektiv empfundenen Alter (Schwall, 2012) oder dem eigenen Körperbild (Jackson et al., 1988). Diese psychologisch-sozialen Einflussfaktoren führen dazu, dass bei der Selbstbewertung der Merkmale eine systematische Über- bzw. Unterschätzung stattfindet. Beim Merkmal Körpergröße spielen derlei Faktoren vermutlich eine untergeordnete Rolle, was sich in einer realistischeren und damit präziseren Selbsteinschätzung niederschlägt. Untermuert wird dies durch Studien zur Diskrepanz zwischen selbstberichteten und objektiv gemessenem Körpergewicht bzw. gemessener Körpergröße (z. B. Larson, 2000; für ein Review vgl. Gorber, Tremblay, Moher & Gorber, 2007). Im Vergleich zu den objektiv erfassten Werten unterliegt die selbstberichtete Körpergröße einer leichten, aber systematischen Überbestimmung. Das selbstberichtete Körpergewicht unterliegt einer systematischen, mit zunehmendem Körpergewicht größeren Unterbestimmung. Die Überbestimmung der selbstberichteten Körpergröße wird auf einen einfachen Fehler zurückgeführt. Erwachsene Menschen unterziehen sich selten Körpergrößensmessungen, weshalb die durch Alter und Schwerkraft entstehende physiologische Schrumpfung von ihnen oft unentdeckt bleibt (Rowland, 1990). Im Gegensatz dazu wird die Unterbestimmung des selbstberichteten Körpergewichts auf soziale Normen zurückgeführt, die einen schlanken Körper propagieren (Ziebland, Thorogood, Fuller & Muir, 1996). Die Güte psychologischer Messungen könnte also eher vom Merkmal selbst als von Faktoren wie formalen Regeln zur Itemformulierung oder dem Skalenformat abhängig sein. Möglicherweise fällt die Merkmalsmessung dann umso genauer aus, je weniger das zu erfassende Merkmal individuellen Einflussfaktoren unterliegt.

Die Items 6 („Ich mache mir viele Gedanken über das Älterwerden.“) und 35 („Wenn ich die Wahl habe zwischen Aufzug und Treppe, dann nehme ich den Aufzug.“) wiesen die schlechtesten psychometrischen Eigenschaften auf. Es zeigte sich kein Zusammenhang zwischen den Items und den realen Werten der Merkmale, die sie erfassen sollten. Möglicherweise erfasst Item 6 *Rumination* im Sinne von Grübeln (Harvey, Watkins, Mansell & Shafran, 2004), da es Ähnlichkeiten zu Items der *Deutschen Version der Skala Rumination aus dem Rumination-Reflection Questionnaire* (König, 2012) aufweist. Item 35 könnte physische Aktivität erfassen, da es Ähnlichkeiten zu Items des *Global Physical Activity Questionnaire* (Armstrong & Bull, 2006) aufweist. Welches Konstrukt die Items statt des

angedachten Merkmals erfassen, lässt sich auf Basis der Befunde nicht abschließend klären. Allerdings konnte gezeigt werden, dass Items mit durchgängig mangelhaften psychometrischen Eigenschaften nicht das intendierte Konstrukt erfassen. Trotz sorgfältiger Itemkonstruktion, kognitiver Interviews und Pilottest konnten derlei Items nicht vollständig vermieden werden. Selbst etablierte Testverfahren, wie das *NEO-PI-R* (Ostendorf & Angleitner, 2004), enthalten nach jahrelanger Entwicklung und Revision Items, die mangelhafte psychometrische Eigenschaften aufweisen. Beispielsweise verfügen die Items 58 aus der Skala Offenheit („Ich meine, dass sich die Gesetze und die Sozialpolitik ändern sollten, um den Bedürfnissen der sich ändernden Welt Rechnung zu tragen.“) und 59 aus der Skala Verträglichkeit („In Bezug auf meine Einstellungen bin ich nüchtern und unnachgiebig.“) über vergleichbare psychometrische Eigenschaften wie die Items 6 und 35 (Trennschärfe Form S für Item 58:  $r_{it} = .12$  und Item 59:  $r_{it} = .11$ ; Ostendorf & Angleitner, 2004, S. 93). Vermutlich erfassen auch diese Items nicht das jeweils angedachte Konstrukt. Für die Fragebogenkonstruktion bedeutet dies, dass derlei Items konsequent ausgeschlossen werden sollten, auch wenn die mangelnde Validität bei psychologischen Konstrukten aufgrund fehlender realer Werte nicht so plakativ gezeigt werden kann wie in Studie III.

Es zeigten sich auffällige positive Zusammenhänge zwischen den Körpergrößenitems und den realen Werten des Merkmals Körpergewicht. Umgekehrt war dieses Muster deutlich weniger ausgeprägt. Folglich transportierten die Körpergrößenitems Informationen, die die Körpergewichtsisems zumindest in dem Ausmaß nicht enthalten. Ein Grund dafür könnte sein, dass ein größerer Körper normalerweise mehr Körpermasse hat, wohingegen mehr Körpermasse nicht zwangsläufig in mehr Körpergröße resultiert. Genau diesem Umstand trägt der BMI Rechnung (Birbaumer & Schmidt, 2010), weshalb beispielsweise bei Amputationen vor Bestimmung des BMI die theoretische Körpermasse mithilfe von Korrekturwerten berechnet werden muss (Tzamaloukas, Patron & Malhotra, 1994). Anhand des BMI lässt sich das Körpergewicht zur Körpergröße relativieren, umgekehrt kann damit nicht die Information über das Körpergewicht von den Körpergrößenitems getrennt werden. Dies spricht gegen Eindimensionalität bei den vorliegenden Körpergrößenitems, womit eine notwendige Voraussetzung für die Gültigkeit der meisten Testmodelle verletzt wäre (Eid & Schmidt, 2014; Rost, 2004). In der Tat wurden alle in Studie III geschätzten Testmodelle

durch den Modelltest abgelehnt. Dennoch verfügen die Körpergrößenitems über die besten psychometrischen Eigenschaften und über die höchste Prädiktionsgüte. Damit wird sichtbar, dass die im Rahmen der Testkonstruktion geforderte Eindimensionalität (z. B. Bühner, 2011; Eid & Schmidt, 2014) selbst bei klar umschriebenen und wenig von individuellen Faktoren beeinflussten Merkmalen, wie Körpergröße, kaum erreichbar ist, ihre Verletzung jedoch nicht zwangsläufig mit einer schlechten Vorhersagegüte einhergehen muss.

Bei Regressionsanalysen auf Basis des BMI konnte in Studie III ein höherer Varianzanteil aufgeklärt werden als bei Regressionsanalysen auf Basis des absoluten Körpergewichts. Die in Studie I angeführte Überlegung, dass eine Korrektur des Körpergewichts um die Körpergröße mithilfe des BMI vor allem bei gemeinsamer Datenauswertung beider Geschlechter sinnvoll ist, scheint damit bestätigt. Allerdings verdeutlicht sich dadurch ein anderes Problem: Laut Definition des zu erfassenden Konstrukts sollten die Items Körpergewicht im Sinne physikalischer Körpermasse erfassen (Martin, 1929). Nachdem der BMI bessere Vorhersagen ermöglichte als das absolute Körpergewicht, erfassten die Items vermutlich eher Körperform als Körpermasse. Somit wird evident, dass selbst eine klare Merkmalsdefinition und sorgfältige Itemkonstruktion zu Items führen können, die nicht hauptsächlich das intendierte Merkmal erfassen. Handelt es sich dabei um psychologische Konstrukte, die so eng miteinander verknüpft sind wie Körpermasse und Körperform, bleibt dies möglicherweise unentdeckt. Damit ginge die psychologische Messung unbemerkt am angedachten Konstrukt vorbei.

Flexiblere probabilistische Testmodelle (2PL-Modell und GPCM), die konzeptuell den tau-kongenerischen Modellen der KTT entsprechen, wiesen eine relativ bessere Modellpassung auf im Vergleich zu restriktiveren Modellen (1PL-Modell und PCM), die konzeptuell den tau-äquivalenten Modellen der KTT entsprechen. Dies geht Hand in Hand mit der Studie von Bortolotti und Kollegen (2013). Die Autoren gingen nicht von gleichen Trennschärfen aus und spezifizierten ein 2PL-Modell. In Studie III musste die globale Modellgültigkeit für alle Modelle verworfen werden. Diese wurde von Bortolotti und Kollegen (2013) nicht überprüft, weshalb hierzu kein Vergleich gezogen werden kann. Unter beiden Bedingungen entsprechen die Stichprobengrößen den für diese probabilistischen Modelle gegebenen Empfehlungen (Bühner, 2011). Es ist nicht zu erwarten, dass bei größeren Stichproben der Modellfit tendenziell besser ausfällt, da die Teststärke der

Modelltests mit zunehmender Stichprobengröße steigt (Eid & Schmidt, 2014). Darum kann die Modellgültigkeit nicht angenommen werden.

Die im Rahmen der CFA spezifizierten Modelle konnten die Daten ebenfalls nicht angemessen beschreiben. Die Gründe für oder gegen einen Modellfit werden in zahlreichen Studien diskutiert (z. B. Browne, MacCallum, Kim, Andersen & Glaser, 2002; Heene, Hilbert, Draxler, Ziegler & Bühner, 2011; Marsh, Balla & McDonald, 1988; Marsh, Hau & Wen, 2004). Beispielsweise werden die Grenzwerte der Fit-Indizes und deren Invarianz gegenüber Ladungshöhen (Heene et al., 2011) oder Stichprobengrößen (Marsh et al., 1988) hinterfragt. Die benötigte Stichprobengröße ist abhängig von verschiedenen Faktoren wie Modellkomplexität, durchschnittlicher Ladungshöhe oder Ausmaß der Verteilungsverletzung. Über diese Faktoren müssen mitunter hypothetische Annahmen getroffen werden, auf Basis derer stark divergierende Empfehlungen zur Mindeststichprobengröße abgeleitet werden (z. B. Kline, 2011; MacCallum, Browne & Sugawara, 1996; Marsh, Hau, Balla & Grayson, 1998; Weston & Gore, 2006). Die Stichprobengrößen dieser Arbeit liegen deutlich im oberen Bereich der gegebenen Empfehlungen. Es ist nicht zu erwarten, dass bei einer noch größeren Stichprobe der Modellfit tendenziell besser ausfallen würde, da die Teststärke der Modelltests mit zunehmender Stichprobengröße steigt (Eid & Schmidt, 2014). Darum kann die Gültigkeit, für die im Rahmen der CFA spezifizierten Modelle nicht angenommen werden.

Unter der Annahme der Modellgeltung werden Vermutungen angestellt, warum sich kein klares Muster in der Frage zeigte, ob Summenwerte, Faktorwerte oder Personenparameter mehr Varianz aufklären. Die den Faktorwerten und Personenparametern zugrundeliegenden Modelle sind für die vorliegende Datenstruktur theoretisch sinnvoll (Eid & Schmidt, 2014). Ihre Schätzung erfolgte an identischen Stichproben. Diskrepanzen aufgrund theoretisch unangemessener Modelle oder unterschiedlicher Stichprobenzusammensetzungen können deshalb ausgeschlossen werden. Die Unterschiede in der Vorhersagegüte zwischen Faktorwerten und Personenparametern sind möglicherweise auf Unterschiede in den Schätzmethoden für Item- und Personenparameter der jeweiligen Modelle zurückzuführen.

Bei Regressionen auf Basis von Summenwerten konnte ein vergleichbarer Anteil an Varianz aufgeklärt werden wie auf Basis der beiden anderen Parameter. Eine Erklärung



dafür könnte das *Verzerrung-Varianz-Dilemma* sein (James, Witten, Hastie & Tibshirani, 2013). Damit ist gemeint, dass die Vorhersagegüte eines Modells von zwei Fehlerquellen abhängt: *Verzerrung* und *Varianz*. Die Verzerrung beschreibt das Ausmaß falscher Annahmen über das Populationsmodell, die Varianz das Ausmaß an Schätzungenauigkeiten des Modells. Flexiblere Modelle können sich besser an den funktionalen Zusammenhang in der Population anpassen, unterliegen jedoch im Mittel größeren Schätzungenauigkeiten als einfachere Modelle. Im Gegensatz dazu sind Modelle mit höherer Verzerrung meist relativ einfach, produzieren aber Vorhersagen mit niedrigerer Varianz. Die modellbasierten Parameter beruhen auf einem relativ flexiblen Modell, die Summenwerte auf einem einfachen Modell. Bei Summenwerten werden die Antworten aller Items aufsummiert. Damit wird angenommen, dass jedes Item einen gleich hohen Vorhersagebeitrag leistet und somit alle Items eine Gewichtung von Eins haben. Die flexibleren Modelle treffen die Annahme, dass Items einen unterschiedlich hohen Beitrag zur Vorhersage des Konstrukts leisten und deshalb unterschiedlich stark gewichtet werden (Eid & Schmidt, 2014). Die Schätzgenauigkeit von Modellen steigt mit zunehmender Stichprobengröße. Möglicherweise sind die vorliegenden Stichprobengrößen nicht ausreichend für eine stabile Schätzung der modellbasierten Parameter. Deshalb könnten sich durch das den Summenwerten zugrundeliegende sparsamere Modell stabilere Schätzungen und damit bessere Vorhersagen ergeben als durch flexiblere Modelle, bei denen Items unterschiedliche, möglicherweise unzutreffende bzw. unsicher geschätzte Gewichtungen erhalten.

Eine Alternativerklärung könnte sein, dass modellbasierte Parameter ein anderes als das angedachte Konstrukt vorhersagen. Möglicherweise wurden durch die modellbasierten Parameter latente Variablen zutreffend geschätzt, die mit den zur Validierung herangezogenen realen Werten nicht stark genug zusammenhängen. Beispielsweise könnten die den Items zugrundeliegenden latenten Variablen nicht chronologisches Alter bzw. Körpergröße sein, sondern subjektiv empfundenes Alter bzw. subjektive Größenwahrnehmung.

#### **4.4.3 Limitationen und Ausblick**

Die Teilnehmerzahl von insgesamt 1 854 Personen ist zufriedenstellend. Die beiden Substichproben sind in ihrer demographischen Verteilung weitestgehend vergleichbar, weshalb Effekte aufgrund zu großer Unterschiedlichkeit vernachlässigt werden können.

Jedoch zeigte sich erneut ein unausgewogenes Verhältnis bei den Merkmalen Bildung (mindestens Abitur oder Fachabitur: 82.3 %) und Alter (linkssteile Altersverteilung;  $Md = 25$ ). Unter Abschnitt 3.4.3 wurde die Abhängigkeit des Antwortverhaltens von Bildungsniveau (Meisenberg & Williams, 2008) und Alter (Weijters et al., 2010) bereits erwähnt. Die Generalisierbarkeit der Ergebnisse ist somit auch in Studie III eingeschränkt. Die Replikation an einer repräsentativen Stichprobe wäre wünschenswert.

Summenwerte, Faktorwerte und Personenparameter wurden an derselben Stichprobe geschätzt, an der sie validiert wurden. Vermutlich führte dieses Vorgehen zu einer Überschätzung der aufgeklärten Varianzanteile  $R^2$ . So kann nicht ausgeschlossen werden, dass die ermittelten Parameter bereits spezifische Informationen über die Stichprobe enthalten, für die sie Vorhersagen treffen sollen. Um einen verlässlicheren Wert für die Prädiktionsgüte der Parameter zu erhalten, müsste man diese an einer neuen, unabhängigen Stichprobe überprüfen. Eine Alternative wäre die Anwendung eines Kreuzvalidierungsverfahrens, beispielsweise der  $k$ -fachen Kreuzvalidierung. Dabei wird die zur Verfügung stehende Stichprobe in  $k$  möglichst gleich große Teilmengen aufgeteilt. Die erste Teilmenge wird zur Validierung verwendet, die Parameterschätzung erfolgt an der verbleibenden Teilmenge  $k-1$ . Dieser Vorgang wird  $k$ -Mal wiederholt, wobei der Validierungssatz immer eine andere Teilmenge ist. Die Schätzung der Vorhersagegüte des Modells erfolgt durch Mittelwertbildung der Werte aus den  $k$  Einzeldurchläufen (James et al., 2013). Bei einer Datenreanalyse sollte diese Möglichkeit mit in Betracht gezogen werden.

Bei der Regressionsanalyse wurde verglichen, ob Faktorwerte aus dem CFA-Modell oder Personenparameter aus dem probabilistischen Testmodell die realen Werte besser vorhersagen können. Bei den CFA-Modellen wurden bei der Parameterschätzung die Faktorwerte für alle Items gemeinsam ermittelt. Bei den probabilistischen Modellen wurden die Personenparameter für jede Skala getrennt ermittelt. Es könnte sein, dass die Faktorwerte insofern einen Vorteil gegenüber den Personenparametern haben, da bei deren Ermittlung Informationen aller Items einfließen. Der vorgenommene Vergleich könnte deshalb unfair sein. Um dies auszuschließen, könnten bei einer Datenreanalyse die Faktorwerte innerhalb eines für jedes Merkmal getrennten CFA-Modells ermittelt und

erneut mit den Personenparametern aus den probabilistischen Modellen verglichen werden.

#### **4.4.4 Fazit**

Zielsetzung von Studie III war die Überprüfung der Frage, wie gut psychologische Skalen in der Lage sind, die physikalisch messbaren Personenmerkmale Alter, Körpergröße und Körpergewicht zu erfassen. Bei einfachen linearen Regressionen mit Summenwerten, Faktorwerten oder Personenparametern als unabhängige Variablen und Alter, Körpergröße oder Körpergewicht bzw. BMI als abhängige Variablen konnte beim Merkmal Körpergröße die meiste Varianz aufgeklärt werden und beim Merkmal Körpergewicht (in Kilogramm) die geringste. Es zeigte sich kein eindeutiges Muster, ob durch Summenwerte, Faktorwerte oder Personenparameter mehr Varianz aufgeklärt werden kann und ob mit einer dichotomen oder polytomen Antwortskala die realen Werte in den Regressionsanalysen besser vorhergesagt werden können.

Das Beispiel Körpergröße zeigt, dass Messungen physikalischer Merkmale mit psychologischen Skalen grundsätzlich möglich sind. Deren Genauigkeit nimmt aber vor allem bei stark von subjektiven Faktoren beeinflussten Merkmalen wie Alter und Körpergewicht ab. Die Güte psychologischer Messungen ist vermutlich eher vom Merkmal selbst abhängig als von Faktoren wie formalen Regeln der Itemformulierung oder dem Skalenformat.

Selbst nach sorgfältiger Skalenkonstruktion konnte nicht ausgeschlossen werden, dass Items in die Skala aufgenommen wurden, die das intendierte Merkmal gar nicht erfassen oder dass Items konstruiert wurden, die hauptsächlich ein anderes, mit dem intendierten Merkmal eng verknüpft Merkmal erfassen.

Möglicherweise wurde durch ein Verzerrung-Varianz-Dilemma sichtbar, dass in der Psychologie die Stichprobengrößen, die für stabile Parameterschätzungen in testtheoretischen Modellen notwendig sind, deutlich unterschätzt werden.

## 5. Abschließende Diskussion

Die übergeordnete Frage der vorliegenden Arbeit war, ob die Psychologie tatsächlich das misst, was sie vorgibt zu messen. In drei empirischen Studien wurden verschiedene Aspekte der Gültigkeit psychologischer Messungen untersucht und welche Einflussfaktoren dabei eine Rolle spielen. In jeder der Studien wurden den psychologischen Messungen reale Werte physikalischer Messgrößen gegenübergestellt. Die Studien I und II untersuchten die psychometrische Güte einzelner Items, die physikalisch messbare Personenmerkmale (Alter, Körpergröße, Körpergewicht und Einkommen) erfassen. Zentrale Fragestellungen der Studien waren, ob Bildungsniveau und relatives Körpergewicht (Studie I) sowie Itemformulierung und Skalenformat (Studie II) Einflussvariablen sind, die dazu führen, dass die Relationen der realen Werte im Mittel unterschiedlich gut auf der Antwortskala erhalten bleiben. Studie III untersuchte den Zusammenhang zwischen physikalisch messbaren Personenmerkmalen (Alter, Körpergröße und Körpergewicht) und deren Abbildung auf psychologische Skalen.

In Studie I wurde den Fragen nachgegangen, ob die Selbstbewertung der Merkmale Alter, Körpergröße, Körpergewicht und Einkommen abhängig vom Bildungsniveau unterschiedlich ausfällt und ob der BMI einen inkrementellen Beitrag für die Vorhersage des Antwortverhaltens bei Selbsteinschätzungen des Gewichts über das absolute Körpergewicht hinaus liefert. Im Mittelpunkt der Analysen stand die Frage, unter welchen Bedingungen sich im Mittel die größte Annäherung an Erhalt von Ordnung, Unterschieden und Verhältnissen der realen Werte auf der Antwortskala ergibt. Bei keinem Merkmal veränderte sich die Stärke des Zusammenhangs zwischen den realen Werten und der Antwortskala abhängig vom Bildungsniveau. Die Regressionsanalysen auf Basis des BMI fielen im Vergleich zu den Regressionsanalysen auf Basis des absoluten Körpergewichts bei Männern vergleichbar und bei Frauen in Bezug auf die Schwellenabstände ungünstiger aus. Innerhalb der Geschlechter konnte der BMI keinen inkrementellen Beitrag für die Vorhersage des Antwortverhaltens bei Selbsteinschätzungen des Gewichts über das absolute Körpergewicht hinaus liefern. Es wurde geschlussfolgert, dass die durch den BMI angedachte Korrektur möglicherweise durch die nach Geschlecht getrennte Datenauswertung bereits vorgenommen wurde und die Korrektur deshalb vor allem bei einer gemeinsamen Datenauswertung beider Geschlechter sinnvoll ist.

In Studie II wurde der Frage nachgegangen, ob die Befragten bei der Selbstbewertung der Merkmale Alter, Körpergröße, Körpergewicht und Einkommen durch Itemformulierung, Skalenverankerung und Skalenbreite im Antwortverhalten in eine bestimmte Richtung beeinflusst werden. So bezogen sich in Studie I die Items durchgehend auf einen Pol des erfragten Merkmals. Die verwendete fünfstufige Likert-Skala (Likert, 1932) war sowohl verbal als auch numerisch verankert. In Studie II wurden Itemformulierung und Antwortskala modifiziert, sodass die Itemformulierung keinen Bezug mehr auf einen Pol des untersuchten Merkmals aufwies und die Antwortskala lediglich in den Endpunkten mit den Worten *niedrig* bzw. *hoch* verbal verankert war. Drei der gängigsten Skalenbreiten (fünfstufiges, vierstufiges und dichotomes Antwortformat) wurden verglichen. Außerdem wurde untersucht, ob sich die Selbstbewertungsperspektive (egozentrisch und allozentrisch) als eine den Referenzrahmen der Selbstbewertung bestimmende Variable zeigt und ob die Selbstbewertung abhängig vom Bildungsniveau unterschiedlich ausfällt. Fokus der Analysen war die Frage, unter welcher Bedingung sich im Mittel die größte Annäherung an Erhalt von Ordnung, Unterschieden und Verhältnissen der realen Werte auf der Antwortskala ergibt (polytome Antwortformate) bzw. inwieweit der Wendepunkt der logistischen Funktion mit den Lagemaßen der zentralen Tendenz korrespondiert (dichotomes Antwortformat). Es gab kein klares Anzeichen für den besseren Erhalt der Relationen der realen Werte bei einer der untersuchten Skalenbreiten. Nur bei den Merkmalen Körpergewicht und Alter blieben bei modifizierter Itemformulierung und Skalenverankerung die Relationen der realen Werte besser erhalten. Jedoch wurde diskutiert, ob die beobachteten Unterschiede eindeutig genug sind, um sie systematischen Einflüssen zuschreiben zu können. Resümee war, dass Itemformulierung, Art der Skalenverankerung und Skalenbreite keinen klar sichtbaren Einfluss auf die Selbstbewertung haben. Durch unterschiedliche Itemformulierung konnte nur bei den Merkmalen Einkommen (Männer) und Körpergewicht (beide Geschlechter) ein Perspektivenwechsel in der Selbstbewertung erreicht werden. Die Selbstbewertungsperspektive kam damit bei Merkmalen zum Tragen, die stark von subjektiven Faktoren beeinflusst werden. Sollen die erhobenen Werte möglichst nah an den realen Werten liegen, ist die allozentrische Formulierungsvariante der egozentrischen vorzuziehen. Für das Merkmal Alter fielen die Parameterschätzungen wie bereits in Eichhorn (2016) äußerst unplausibel aus (z. B. geschätzte Schwellenparameter bei einem Alter über 400 Jahren). Es wurde diskutiert, dass dieses Merkmal aufgrund seiner Komplexität und

Mehrdimensionalität besonders messfehleranfällig ist. Die Ergebnisse beim Merkmal Alter sollten deshalb mit Vorsicht interpretiert werden. Die Selbstbewertung blieb unabhängig vom Bildungsniveau, womit die Ergebnisse aus Studie I bestätigt werden konnten.

In Studie III wurde der Zusammenhang zwischen den Merkmalen Alter, Körpergröße und Körpergewicht und deren Abbildung auf psychologische Skalen untersucht. Die Skalenkonstruktion für die Merkmale folgte den in der Fachliteratur gegebenen Empfehlungen zur Fragebogenkonstruktion und umfasste Experteninterviews, einen Prototypenansatz, kognitive Interviews und einen Pilottest. Die Beantwortung der Items erfolgte entweder auf einer polytomen (sechsstufig) oder dichotomen Antwortskala. Im Mittelpunkt der Analysen stand die Frage, wie gut die Skalen in der Lage sind, die physikalischen Merkmale abzubilden. Für beide Antwortskalen ergaben sich beim Merkmal Körpergröße die höchsten Reliabilitätsschätzungen und die größten aufgeklärten Varianzanteile. Es zeigte sich kein eindeutiges Muster, ob mit Summenwerten, Faktorwerten oder Personenparametern mehr Varianz aufgeklärt werden kann. Der Vergleich beider Antwortskalenformate konnte keine eindeutige Aussage darüber liefern, welches der Skalenformate die realen Werte in den Regressionsanalysen besser vorhersagen kann. Unter keiner Bedingung und bei keinem Merkmal konnte eine Gültigkeit der spezifizierten Modelle angenommen werden. Es wurde resümiert, dass Messungen physikalischer Merkmale mit psychologischen Skalen grundsätzlich möglich sind, die Messgenauigkeit aber vor allem bei stark von subjektiven Faktoren beeinflussten Merkmalen abnimmt. Die Güte psychologischer Messungen würde damit eher vom Merkmal selbst abhängen als von Faktoren wie Itemformulierung oder Skalenformat.

Wie das Merkmal Körpergröße zeigt, ist die Durchführung von Messungen mittels Items grundsätzlich keine schlechte Herangehensweise. So blieben beim Merkmal Körpergröße die Relationen der realen Werte am besten auf den Antwortskalen erhalten (Studie I und II). Verglichen mit den anderen Merkmalen konnte die meiste Varianz aufgeklärt werden (Studie III). Fraglich ist jedoch, ob das Merkmal Körpergröße repräsentativ für typische psychologische Konstrukte ist. Von den untersuchten Merkmalen scheint es das Merkmal mit der geringsten Komplexität zu sein. Vermutlich transportiert es wenig psychologische Bedeutung und unterliegt nur in geringem Ausmaß subjektiven Einflussfaktoren.

Im Gegensatz dazu weisen vor allem die Merkmale Alter (Schwall, 2012) und Körpergewicht (Jackson et al., 1988) eine höhere Vielschichtigkeit auf, womit sie psychologischen Konstrukten näherkommen. So sind diese überwiegend durch Komplexität und Mehrdimensionalität gekennzeichnet, wie beispielsweise Neurotizismus oder Extraversion (Neyer & Asendorpf, 2018). Der Fehlervarianzanteil bewegte sich bei den Merkmalen Alter und Körpergewicht zwischen 39 % und 59 % (vgl. Tabelle 4.3.9.), was in etwa dem Messfehleranteil psychologischer Messungen entspricht. Selbst bei gut etablierten Konstrukten wie den Big Five wird dieser Anteil auf bis zu 50 % geschätzt (Gnambs, 2015). Auch wenn die Persönlichkeitsforschung gelegentlich wegen ungenauer Messungen und ihrer mangelnden prädiktiven Validität kritisiert wurde (Blinkhorn & Johnson, 1990), scheint die hohe Messungenauigkeit in der Psychologie kein Thema von Interesse zu sein. So führte die Kritik nicht zu nennenswerten Neuerungen im Testaufbau oder Bemühungen zur Messfehlerreduktion (Gnambs, 2015). Angesichts der Replikationskrise, die darauf hinweist, dass viele psychologische Entdeckungen auf wenig belastbaren Beweisen beruhen (Klein et al., 2014a; Klein et al., 2014b; Klein et al., 2015; Open Science Collaboration, 2015), wäre die Psychologie jedoch gut beraten, mehr Anstrengungen zur Verbesserung psychologischer Messungen zu unternehmen.

Es hat den Anschein, dass die Entwicklung der aktuellen Messtheorie primär mathematisch-formalen Aspekten folgt und sich von der Empirie wegentwickelt. Bereits Stevens (1946) definierte seine Skalenniveaus auf Basis mathematischer Transformationen. Empirische Überlegungen kamen dabei an keiner Stelle wirklich zum Tragen. Auch das Festhalten am Homomorphismus hat seine Problematik (Saint-Mont, 2011). Mag er innerhalb der Mathematik selbstverständlich sein, stellt er an Messungen fundamentale Anforderungen. Die Voraussetzung, dass Relationen zwischen Zahlen *genau dann* gelten, wenn sie auch zwischen den Objekten in der Realität bestehen, grenzt an Naivität. So bleiben zu erwartende Messfehler generell unberücksichtigt. Stattdessen wird das perfekte Korrelat numerischer und empirischer Strukturen modelliert, welches keine Fehler vorsieht. Dieses Modell setzt damit voraus, dass die gemessene Welt ebenso exakt und fehlerfrei ist und dass man in der Lage ist, sie auch als solche erkennen zu können. Nachdem diese notwendige Voraussetzung in der Realität kaum gegeben sein dürfte, ist die Gültigkeit dieses theoretischen Modells fraglich (Saint-Mont, 2011).

In der Einleitung wurde die Wichtigkeit der Mathematisierbarkeit psychischer Phänomene für die (natur-)wissenschaftliche Anerkennung der Psychologie verdeutlicht. Erklärtes Ziel sollte dabei jedoch sein, dass die Mathematisierung einem „kristallisierten Verständnis realer Sachverhalte entspricht“ (Saint-Mont, 2011, S. 43). Betrachtet man die aktuelle Entwicklung der psychologischen Messtheorie, ist die Erfüllung dieses Anspruchs zweifelhaft. Vielmehr entsteht der Eindruck, dass es um Abstraktion der Abstraktion willen geht, wobei der Bezug zur konkreten Problemstellung zunehmend verloren geht (Saint-Mont, 2011). Schlägt man neuere Werke zur Test- und Fragebogenkonstruktion auf, hält man im Wesentlichen ein Mathematikbuch in Händen. Praktisch relevante Anleitungen zur Ausführung konkreter Messvorgänge fehlen häufig. Selbst die Richtlinien zur Itemformulierung sind oft nicht verallgemeinerbare, sehr punktuelle oder schlicht banale Empfehlungen. Obwohl derlei Hinweise ihre Berechtigung und ihren Sinn haben, ist ihr Nutzen beschränkt. Mehr als grobe Orientierung bei der Itemformulierung können sie nicht geben (Porst, 2014).

Wie unter Abschnitt 4.1.3 beschrieben, liegt psychologischen Fragebögen die theoretische Annahme einer latenten Variablen zugrunde. Diese Annahme wird innerhalb statistischer Modelle mathematisch formalisiert und überprüft, beispielsweise im Rahmen einer CFA (Bühner, 2011). Wie auch in dieser Arbeit, ist die Gültigkeit der meisten Modelle nicht gegeben (z. B. Heene et al., 2011; Marsh et al., 1988; Marsh et al., 2004; Pargent et al., 2018). Infolgedessen werden immer komplexere und anspruchsvollere statistische Modelle entwickelt. Das Fortschreiten in höhere Komplexitätsstufen führte jedoch weder zu einer besseren Modellgeltung noch zu einer besseren Abbildung der Empirie (z. B. Pargent et al., 2018). Berechtigterweise stellt sich also die Frage, was der statistische Nutzen dieser Modelle ist, wenn sie sich im konkreten Anwendungsfall für „nicht zuständig“ erklären (Saint-Mont, 2011, S. 49). Die angenommene Beziehung zwischen Theorie und Empirie lässt sich nicht zeigen, womit die mathematische Lösung eine Scheinlösung ist. Rein formale Zusammenhänge stehen auf der einen und die empirische Bedeutung auf der anderen Seite (Saint-Mont, 2011). Dabei müssen sich jede Theorie und jedes Modell an der Wirklichkeit messen lassen. So ist nach Tukey (1962, S. 5) „reliance upon the test of experience as the ultimate standard of validity“ für alle wissenschaftlichen Theorien entscheidend.



Sowohl messtheoretische Grundlagen als auch die Empirie haben ihre Berechtigung. Ein Kompromiss, der beiden Seiten gerecht wird, wäre angemessen, ist jedoch nicht leicht zu finden. Zur Entschärfung des Problems werden deshalb vor allem begriffliche Differenzierungen vorgenommen, ohne dabei die Anwendbarkeit der ganzen Theorie in Frage zu stellen (für einen Überblick vgl. Niederée, 1994; Steyer & Eid, 1993). Ein klassischer Ansatz ist die Unterscheidung zwischen *fundamentalen* und *sonstigen* Messungen (Adams, Fagot & Robinson, 1965). Bei fundamentalen Messungen ist eine direkte Zuordnung von Zahlen zu Objekten möglich, ein empirisches Relativ zweifelsfrei erkennbar und damit die messtheoretische Annahme anwendbar. Beispiele dafür sind Längen-, Massen- oder Volumenmessungen (Steyer & Eid, 1993). Bei sonstigen Messungen, innerhalb derer oft nicht klar genug ist, was überhaupt gemessen wird, ist die Messtheorie nicht anwendbar, was nach Adams und Kollegen (1965) aber keinen Makel darstellt.

Die Psychologie kann mit den Präzisionsmessungen von Naturwissenschaften wie der Physik nicht konkurrieren. Dennoch würde man die quantitativen Bemühungen der letzten Jahrzehnte diskreditieren, dürfte die Psychologie nicht von Messungen sprechen. Was die Psychologie anerkennen muss, ist, dass qualitativ hochwertige Messungen nicht mühelos erreichbar sind. Nicht jede Zuordnung von Ziffern zu empirischen Sachverhalten ist eine Messung (Saint-Mont, 2011). Zur Verbesserung psychologischer Messungen müssten deren qualitative Aspekte mehr in den Fokus gerückt werden. Beispielsweise sollte intensiver untersucht werden, welche Eigenschaften das Messinstrument hat, welche Einflussvariablen bei der Messung wirken, wie der Messgegenstand überhaupt definiert ist, welche Fehlerquellen vorhanden sind und wie diese reduziert werden können.

Messungen via Fragebögen nehmen immer den Umweg über die Sprache. Beim Lesen eines Items findet bereits ein Konstruktionsprozess statt, der nicht zwangsläufig bei jeder Person zum selben Ergebnis führt. So müssen zunächst Wörter und syntaktische Strukturhinweise erkannt und in begleitende, außersprachliche Wahrnehmungen eingeordnet werden. Vor dem jeweiligen Kontext- und Weltwissen muss das Geschriebene verstanden werden (Dietrich & Gerwien, 2017). Dabei existieren zahlreiche Fehlerquellen, beispielsweise die semantische Unschärfe oder Mehrdeutigkeit von Begriffen (Wahlster, 1977). Dadurch können, je nach individuellem Erfahrungshintergrund, Begriffen andere Bedeutungen zugewiesen werden (Schwarz, 2007). Infolgedessen ändert sich das

Itemverständnis und damit dessen Beantwortung. Es gilt die Heisenbergsche Unschärferelation, bei der die Messung als ein modellabhängiger und theoriegeleiteter Eingriff in ein System verstanden wird. Die empirische Präzisierung kann nur um den Preis zunehmender Unschärfe anderer Variablen erfolgen. Es ist unmöglich, eine Eigenschaft exakt zu bestimmen, da das Messgerät mit dem Messgegenstand interagiert. Beide sind voneinander abhängig (Baker, 2015; Orth, 1974).

Überdies wäre mehr Wissen über die der Itembeantwortung zugrundeliegenden kognitiven Prozesse wünschenswert. Dadurch könnten die Einflussfaktoren auf den Selbstbewertungsprozess besser verstanden und sinnvoller miteinander in Zusammenhang gebracht werden. Zwar wurde mithilfe allgemeiner Modelle bereits oft versucht, den Prozess der Itembeantwortung schematisch abzubilden (z. B. Krosnick, 1999; für einen Überblick vgl. Tourangeau, Rips & Rasinski, 2000), jedoch bleibt die Prädiktionsgüte dieser Modelle unbefriedigend. Da bei der Erforschung derart komplexer und vielschichtiger kognitiver Prozesse quantitative Methoden ungeeignet sind (Bühner, 2011), wäre zunächst ein qualitativer Zugang sinnvoll. Beispielsweise könnten mit der Technik des lauten Denkens (Ericsson & Simon, 1980) Erkenntnisse gewonnen werden, welche kognitiven Prozesse bei der Selbsteinschätzung eine Rolle spielen und wie diese den Entscheidungsprozess beeinflussen. Fraglich ist, in welchem Ausmaß diese Abläufe den Befragten überhaupt bewusst zugänglich sind. So ist hinreichend bekannt, dass Fragebögen Antwortverzerrungen unterliegen, wie beispielsweise sozialer Erwünschtheit (Mummendey & Grau, 2014). Damit wird die Tatsache beschrieben, dass Personen gewöhnlich danach streben, in einem positiven Licht gesehen zu werden und dahingehend ihre Antworten gestalten (Manstead & Livingstone, 2014). Nach Paulhus (1984) handelt es sich bei sozialer Erwünschtheit um ein zweidimensionales Konstrukt: Selbsttäuschung und Fremdtäuschung. Unter Selbsttäuschung versteht er die *unbewusste* Tendenz, sich selbst in einem besseren Licht zu sehen. Unter Fremdtäuschung betrachtet er eine *absichtlich* herbeigeführte Manipulation, um einen positiven Eindruck bei anderen hervorzurufen.

Nicht zuletzt wäre eine wichtige Voraussetzung für eine qualitativ hochwertige Messung eine eindeutige, vollständige und verbindliche Definition des Messgegenstandes. In Anlehnung an die Psychiatrie, die ihre Phänomene verbindlich in der fünften Auflage des *Diagnostic and Statistical Manual of Mental Disorders* (American Psychiatric Association,

2013) beschreibt und klassifiziert, könnte ein Regelwerk für die wichtigsten psychologischen Konstrukte erstellt werden. Auf Basis dieser allgemeingültigen Festlegungen könnten sodann psychologische Konstrukte und Phänomene erforscht und auf ihre empirische Realität überprüft werden.

Eine weitere zentrale Frage ist zudem, weshalb sich die akademische Psychologie neben den mathematischen nicht auch mit den qualitativen Aspekten des Messens intensiv auseinandersetzt. Möglicherweise besteht die unbewusste Befürchtung, dass Anstrengungen in diesem Bereich nicht zum Erfolg führen. Beispielsweise könnte es sein, dass die der Itembeantwortung zugrundeliegenden kognitiven Prozesse nicht ausreichend geklärt werden können und/oder sich eine Verständigung auf eine Definition des Messgegenstands nicht erzielen lässt. Das Verhalten der akademischen Psychologie kommt damit fast einer institutionalisierten Abwehr gleich (Mentzos, 2017). Vor dem Hintergrund ihrer in der Einleitung geschilderten Geschichte wird dies verständlich. So ist ihre Existenz letztendlich an die Messbarkeit psychologischer Phänomene geknüpft.

Zur Diskussion, ob die Psychologie das misst, was sie vorgibt zu messen, leistet die vorliegende Arbeit wichtige Beiträge. Sie legt nahe, dass die Psychologie mit fundamentalen Messungen zeitgenössischer Naturwissenschaften nicht mithalten kann. Bei einer liberalen Auffassung im Sinne sonstiger Messungen (Adams et al., 1965) sind Messungen via Fragebögen grundsätzlich möglich. Jedoch sind sie, je nach Merkmal, mehr oder weniger hoch mit Messfehlern behaftet. Im Angesicht der Replikationskrise wäre die Psychologie gut beraten, ihre Messinstrumente nicht nur im statistisch-mathematischen Sinne weiterzuentwickeln. Vielmehr sollte sie sich mit dem Messgegenstand als solchem, den an der Messung beteiligten Prozessen sowie den in erheblichem Maß vorhandenen Messfehlern auseinandersetzen. Eine Psychologie, die ihre angenommenen Konstrukte valide messen kann, scheint nicht in unmittelbarer Reichweite.

## Literaturverzeichnis

- Adams, E. W., Fagot, R. F. & Robinson, R. E. (1965). A theory of appropriate statistics. *Psychometrika*, 30(2), 99–127. doi:10.1007/BF02289443
- Agresti, A. (2010). *Analysis of ordinal categorical data* (2nd ed.). Hoboken, NJ: Wiley.
- Agresti, A. (2013). *Categorical data analysis* (3rd ed.). Hoboken, NJ: Wiley.
- Ahima, R. S. & Lazar, M. A. (2013). The health risk of obesity – better metrics imperative. *Science*, 341(6148), 856–858. doi:10.1126/science.1241244
- Aiken, L. R. & Groth-Marnat, G. (2006). *Psychological testing and assessment* (12th ed.). Boston, MA: Pearson.
- American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Washington, DC: Author.
- Anastasi, A. (1982). *Psychological testing* (5th ed.). New York, NY: MacMillan.
- Angleitner, A., John, O. P. & Löhr, F.-J. (1986). It's what you ask and how you ask it: An itemmetric analysis of personality questionnaires. In A. Angleitner & J. S. Wiggins (Eds.), *Personality assessment via questionnaires: Current issues in theory and measurement* (pp. 61–108). Berlin: Springer.
- Armstrong, T. & Bull, F. (2006). Development of the World Health Organization Global Physical Activity Questionnaire (GPAQ). *Journal of Public Health*, 14(2), 66–70. doi:10.1007/s10389-006-0024-x
- Baker, B. O., Hardyck, C. D. & Petrinovich, L. F. (1966). Weak measurements vs. strong statistics: An empirical critique of S. S. Stevens' proscriptions on statistics. *Educational and Psychological Measurement*, 26(2), 291–309. doi:10.1177/001316446602600204
- Baker, J. (2015). *50 Schlüsselideen Quantenphysik*. Berlin: Springer Spektrum. doi:10.1007/978-3-662-45033-8
- Barnette, J. J. (2000). Effects of stem and Likert response option reversals on survey internal consistency: If you feel the need, there is a better alternative to using those negatively worded stems. *Educational and Psychological Measurement*, 60(3), 361–370. doi:10.1177/00131640021970592

- Barrett, A. E. (2005). Gendered experiences in midlife: Implications for age identity. *Journal of Aging Studies*, 19(2), 163–183. doi:10.1016/j.jaging.2004.05.002
- Bassili, J. N. & Scott, B. S. (1996). Response latency as a signal to question problems in survey research. *Public Opinion Quarterly*, 60(3), 390–399. doi:10.1086/297760
- Baumgartner, H. & Steenkamp, J.-B. E. M. (2001). Response styles in marketing research: A cross-national investigation. *Journal of Marketing Research*, 38(2), 143–156. doi:10.1509/jmkr.38.2.143.18840
- Beauducel, A. & Wittmann, W. W. (2005). Simulation study on fit indexes in CFA based on data with slightly distorted simple structure. *Structural Equation Modeling: A Multidisciplinary Journal*, 12(1), 41–75. doi:10.1207/s15328007sem1201\_3
- Beierlein, C., Kovaleva, A., László, Z., Kemper, C. J. & Rammstedt, B. (2014). Eine Single-Item-Skala zur Erfassung der Allgemeinen Lebenszufriedenheit: Die Kurzskala Lebenszufriedenheit-1 (L-1). *GESIS Working Papers* 33, 1–25.
- Ben-Porath, Y. S. & Tellegen, A. (2008/2011). *The Minnesota Multiphasic Personality Inventory-2 Restructured Form (MMPI-2-RF): Manual for administration, scoring, and interpretation*. Minneapolis, MN: University of Minnesota Press.
- Benson, J. & Hocevar, D. (1985). The impact of item phrasing on the validity of attitude scales for elementary school children. *Journal of Educational Measurement*, 22(3), 231–240. doi:10.1111/j.1745-3984.1985.tb01061.x
- Bernaards, C. A. & Jennrich, R. I. (2005) Gradient projection algorithms and software for arbitrary rotation criteria in factor analysis. *Educational and Psychological Measurement*, 65, 676–696 [Computersoftware]. Verfügbar unter <<http://www.stat.ucla.edu/research/gpa>>
- Berthoz, A. (1991). Reference frames for the perception and control of movement. In J. Paillard (Ed.), *Brain and space* (pp. 81–111). New York, NY: Oxford University Press.
- Birbaumer, N. & Schmidt, R. F. (2010). *Biologische Psychologie* (7. Aufl.). Heidelberg: Springer. doi:10.1007/978-3-540-95938-0
- Birkett, N. J. (1986). Selecting the number of response categories for a Likert-type scale. *Proceedings of the American Statistical Association, Section on Survey Research Methods*, 488-492.

- Bissantz, N. & Butterwegge, G. (2013). Das Diagramm ist tot, es lebe das Diagramm. In W. Weber, M. Burmester & R. Tille (Hrsg.), *Interaktive Infografiken* (S. 135–146). Berlin: Springer. doi:10.1007/978-3-642-15453-9
- Blinkhorn, S. & Johnson, C. (1990). The insignificance of personality testing. *Nature*, 348(6303), 671–672. doi:10.1038/348671a0
- Bogner, A., Littig, B. & Menz, W. (2014). *Interviews mit Experten: Eine praxisorientierte Einführung*. Wiesbaden: Springer. doi:10.1007/978-3-531-19416-5
- Borkenau, P. & Ostendorf, F. (2008). *NEO-FFI. NEO-Fünf-Faktoren-Inventar nach Costa und McCrae* (2. Aufl.). Göttingen: Hogrefe.
- Borsboom, D., Mellenbergh, G. J. & van Heerden, J. (2004). The concept of validity. *Psychological Review*, 111(4), 1061–1071. doi:10.1037/0033-295X.111.4.1061
- Bortolotti, S. L. V., Tezza, R., Andrade, D. F. de, Bornia, A. C. & Sousa Júnior, A. F. de. (2013). Relevance and advantages of using the item response theory. *Quality & Quantity*, 47(4), 2341–2360. doi:10.1007/s11135-012-9684-5
- Bortz, J. & Schuster, C. (2010). *Statistik für Human- und Sozialwissenschaftler* (7. Aufl.). Berlin: Springer. doi:10.1007/978-3-642-12770-0
- Bosch, K. (2011). *Elementare Einführung in die Wahrscheinlichkeitsrechnung* (11. Aufl.). Wiesbaden: Vieweg + Teubner.
- Bosco, F. A., Aguinis, H., Singh, K., Field, J. G. & Pierce, C. A. (2015). Correlational effect size benchmarks. *Journal of Applied Psychology*, 100(2), 431–449. doi:10.1037/a0038047
- Bowman, M. L. (1989). Testing individual differences in ancient China. *American Psychologist*, 44(3), 576–578. doi:10.1037/0003-066X.44.3.576.b
- Bradburn, N. M., Sudman, S. & Wansink, B. (2004). *Asking questions: The definitive guide to questionnaire design – for market research, political polls, and social and health questionnaires* (Rev. ed.). San Francisco, CA: Jossey-Bass.
- Brant, R. (1990). Assessing proportionality in the proportional odds model for ordinal logistic regression. *Biometrics*, 46(4), 1171–1178. doi:10.2307/2532457
- Brewer, B. & Pears, J. (1999). Frames of reference. In N. Eilan, R. McCarthy & B. Brewer (Eds.), *Spatial representation: Problems in philosophy and psychology* (pp. 25–30). New York, NY: Oxford University Press.

- Broughton, R. (1984). A prototype strategy for construction of personality scales. *Journal of Personality and Social Psychology*, 47(6), 1334–1346. doi:10.1037/0022-3514.47.6.1334
- Browne, M. W., MacCallum, R. C., Kim, C.-T., Andersen, B. L. & Glaser, R. (2002). When fit indices and residuals are incompatible. *Psychological Methods*, 7(4), 403–421. doi:10.1037//1082-989X.7.4.403
- Brückler, F. M. (2018). *Geschichte der Mathematik kompakt: Das Wichtigste aus Analysis, Wahrscheinlichkeitstheorie, angewandter Mathematik, Topologie und Mengenlehre*. Berlin: Springer. doi:10.1007/978-3-662-55574-3
- Bryman, A. (2016). *Social research methods* (5th ed.). Oxford: Oxford University Press.
- Büchter, A. & Henn, H.-W. (2007). *Elementare Stochastik: Eine Einführung in die Mathematik der Daten und des Zufalls* (2. Aufl.). Berlin: Springer. doi:10.1007/978-3-540-45382-6
- Bühner, M. (2011). *Einführung in die Test- und Fragebogenkonstruktion* (3. Aufl.). München: Pearson.
- Bühner, M. & Ziegler, M. (2017). *Statistik für Psychologen und Sozialwissenschaftler* (2. Aufl.). Hallbergmoos: Pearson.
- Cantor, N. & Mischel, W. (1977). Traits as prototypes: Effects on recognition memory. *Journal of Personality and Social Psychology*, 35(1), 38–48. doi:10.1037/0022-3514.35.1.38
- Cho, E. (2016). Making reliability reliable: A systematic approach to reliability coefficients. *Organizational Research Methods*, 19(4), 651–682. doi:10.1177/1094428116656239
- Christensen, R. H. B. (2015). ordinal: Regression models for ordinal data. R package (Version 2015.6-28) [Computersoftware]. Verfügbar unter <http://www.cran.r-project.org/package=ordinal/>
- Cicchetti, D. V., Showalter, D. & Tyrer, P. J. (1985). The effect of number of rating scale categories on levels of interrater reliability: A Monte Carlo investigation. *Applied Psychological Measurement*, 9(1), 31–36. doi:10.1177/014662168500900103
- Clark, L. A. & Watson, D. (1995). Constructing validity: Basic issues in objective scale development. *Psychological Assessment*, 7(3), 309–319. doi:10.1037/1040-3590.7.3.309

- Cleveland, J. N. & Shore, L. M. (1992). Self- and supervisory perspectives on age and work attitudes and performance. *Journal of Applied Psychology*, 77(4), 469–484.  
doi:10.1037/0021-9010.77.4.469
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.
- Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112(1), 155–159.  
doi:10.1037/0033-2909.112.1.155
- Collani, G. von & Herzberg, P. Y. (2003). Zur internen Struktur des globalen Selbstwertgefühls nach Rosenberg. *Zeitschrift für Differentielle und Diagnostische Psychologie*, 24(1), 9–22. doi:10.1024//0170-1789.24.1.9
- Cordery, J. L. & Sevastos, P. P. (1993). Responses to the original and revised Job Diagnostic Survey: Is education a factor in responses to negatively worded items? *Journal of Applied Psychology*, 78(1), 141–143. doi:10.1037/0021-9010.78.1.141
- Corson, K., Gerrity, M. S. & Dobscha, S. K. (2004). Screening for depression and suicidality in a VA primary care setting: 2 items are better than 1 item. *The American Journal of Managed Care*, 10(11), 839–845.
- Couch, A. & Keniston, K. (1960). Yeasayers and naysayers: Agreeing response set as a personality variable. *The Journal of Abnormal and Social Psychology*, 60(2), 151–174.  
doi:10.1037/h0040372
- Cox III, E. P. (1980). The optimal number of response alternatives for a scale: A review. *Journal of Marketing Research*, 17(4), 407–422. doi:10.2307/3150495
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16(3), 297–334. doi:10.1007/BF02310555
- Cronbach, L. J. & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, 52(4), 281–302. doi:10.1037/h0040957
- Culbertson, H. M. & Powers, R. D. (1959). A study of graph comprehension difficulties. *Audiovisual Communication Review*, 7(3), 97–110.
- Curcio, F. R. (1987). Comprehension of mathematical relationships expressed in graphs. *Journal for Research in Mathematics Education*, 18(5), 382–393. doi:10.2307/749086



- Deister, A. (2009). Persönlichkeitsstörungen. In H.-J. Möller, G. Laux & A. Deister (Hrsg.), *Psychiatrie und Psychotherapie* (4. Aufl., S. 355–375). Stuttgart: Thieme.
- Diers, C. J. (1964). Social desirability and acquiescence in response to personality items. *Journal of Consulting Psychology*, 28(1), 71–77. doi:10.1037/h0043753
- Dieterich, M. (2012). Vestibuläres System und Störungen der vestibulären Raumorientierung. In H.-O. Karnath & P. Thier (Hrsg.), *Kognitive Neurowissenschaften* (3. Aufl., S. 263–272). Berlin: Springer. doi:10.1007/978-3-642-25527-4\_24
- Dietrich, R. & Gerwien, J. (2017). *Psycholinguistik: Eine Einführung* (3. Aufl.). Stuttgart: Springer. doi:10.1007/978-3-476-05494-4
- Döring, N. & Bortz, J. (2016). *Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften* (5. Aufl.). Berlin: Springer. doi:10.1007/978-3-642-41089-5
- Drake, K. M., Hargraves, J. L., Lloyd, S., Gallagher, P. M. & Cleary, P. D. (2014). The effect of response scale, administration mode, and format on responses to the CAHPS clinician and group survey. *Health Services Research*, 49(4), 1387-1399. doi:10.1111/1475-6773.12160
- Drolet, A. L. & Morrison, D. G. (2001). Do we really need multiple-item measures in service research? *Journal of Service Research*, 3(3), 196–204. doi:10.1177/109467050133001
- Eckardt, G. (2010). *Kernprobleme in der Geschichte der Psychologie*. Wiesbaden: VS.
- Eichhorn, K. (2016). *Experimentelle Untersuchung des Skalenniveaus 5-stufiger Likert-Skalen bei Selbstbewertungen* (unveröffentlichte Masterarbeit). Ludwig-Maximilians-Universität, München.
- Eid, M., Gollwitzer, M. & Schmitt, M. (2015). *Statistik und Forschungsmethoden* (4. Aufl.). Weinheim: Beltz.
- Eid, M. & Schmidt, K. (2014). *Testtheorie und Testkonstruktion*. Göttingen: Hogrefe.
- Ellis, P. D. (2010). *The essential guide to effect sizes: Statistical power, meta-analysis, and the interpretation of research results*. Cambridge: Cambridge University Press.
- Epskamp, S. & Stuber, S. (2017). semPlot: Path diagrams and visual analysis of various SEM packages' output. R package (Version 1.1.) [Computersoftware]. Verfügbar unter <https://CRAN.R-project.org/package=semPlot>

- Ericsson, K. A. & Simon, H. A. (1980). Verbal reports as data. *Psychological Review*, 87(3), 215–251. doi:10.1037/0033-295X.87.3.215
- Faulbaum, F., Prüfer, P. & Rexroth, M. (2009). *Was ist eine gute Frage? Die systematische Evaluation der Fragenqualität*. Wiesbaden: VS.
- Finn, J. A., Ben-Porath, Y. S. & Tellegen, A. (2015). Dichotomous versus polytomous response options in psychopathology assessment: Method or meaningful variance? *Psychological Assessment*, 27(1), 184–193. doi:10.1037/pas0000044
- Finn, R. H. (1972). Effects of some variations in rating scale characteristics on the means and reliabilities of ratings. *Educational and Psychological Measurement*, 32(2), 255–265. doi:10.1177/001316447203200203
- Finucane, M. L., Alhakami, A., Slovic, P. & Johnson, S. M. (2000). The affect heuristic in judgments of risks and benefits. *Journal of Behavioral Decision Making*, 13(1), 1–17. doi:10.1002/(SICI)1099-0771(200001/03)13:1<1::AID-BDM333>3.0.CO;2-S
- Fischer, G. H. (1974). *Einführung in die Theorie psychologischer Tests: Grundlagen und Anwendungen*. Bern: Huber.
- Friel, S. N., Curcio, F. R. & Bright, G. W. (2001). Making sense of graphs: Critical factors influencing comprehension and instructional implications. *Journal for Research in Mathematics Education*, 32(2), 124–158. doi:10.2307/749671
- Friendly, M. & Denis, D. J. (2001) *Milestones in the history of thematic cartography, statistical graphics, and data visualization: An illustrated chronology of innovations*. Zugriff am 10.08.2017. Verfügbar unter <http://www.datavis.ca/milestones/>
- Gaito, J. (1960). Scale classification and statistics. *Psychological Review*, 67(4), 277–278. doi:10.1037/h0043043
- Gardner, P. L. (1975). Scales and statistics. *Review of Educational Research*, 45(1), 43–57. doi:10.3102/00346543045001043
- Garland, R. (1991). The mid-point on a rating scale: Is it desirable. *Marketing Bulletin*, 2(1), 66–70.
- Gerlitz, J. Y. & Schupp, J. (2005). Zur Erhebung der Big-Five-basierten Persönlichkeitsmerkmale im SOEP. Dokumentation der Instrumentenentwicklung BFI-S auf Basis des SOEP-Pretests 2005. *DIW Research Notes*, 4.

- GESIS - Leibniz-Institut für Sozialwissenschaften (2015). *ALLBUS/GGSS 2014: Allgemeine Bevölkerungsumfrage der Sozialwissenschaften/German General Social Survey 2014* (GESIS Data Archive, Cologne. ZA5240 Data file Version 2.1.0) [Datendatei].  
doi:10.4232/1.12288
- Giancoli, D. C. (2010). *Physik: Lehr- und Übungsbuch* (3. Aufl.). München: Pearson.
- Glazer, N. (2011). Challenges with graph interpretation: A review of the literature. *Studies in Science Education*, 47(2), 183–210. doi:10.1080/03057267.2011.605307
- Gnambs, T. (2015). Facets of measurement error for scores of the Big Five: Three reliability generalizations. *Personality and Individual Differences*, 84, 84–89.  
doi:10.1016/j.paid.2014.08.019
- Gorber, S. C., Tremblay, M., Moher, D. & Gorber, B. (2007). A comparison of direct vs. self-report measures for assessing height, weight and body mass index: A systematic review. *Obesity Reviews*, 8(4), 307–326. doi:10.1111/j.1467-789X.2007.00347.x
- Goretzko, D., Pargent, F., Sust, L. N. N. & Bühner, M. (2018). *Not very powerful – the influence of negations and vague quantifiers on psychometric properties of questionnaire items*. Manuscript under review.
- Gortmaker, S. L., Must, A., Perrin, J. M., Sobol, A. M. & Dietz, W. H. (1993). Social and economic consequences of overweight in adolescence and young adulthood. *The New England Journal of Medicine*, 329(14), 1008–1012.  
doi:10.1056/NEJM199309303291406
- Graumann, C. F. (1980). Wundt vor Leipzig – Entwürfe einer Psychologie. In W. Meischner & A. Metge (Hrsg.), *Wilhelm Wundt – Progressives Erbe, Wissenschaftsentwicklung und Gegenwart* (S. 63–77). Leipzig: Karl-Marx-Universität.
- Greenberger, E., Chen, C., Dmitrieva, J. & Farruggia, S. P. (2003). Item-wording and the dimensionality of the Rosenberg Self-Esteem Scale: Do they matter? *Personality and Individual Differences*, 35(6), 1241–1254. doi:10.1016/S0191-8869(02)00331-8
- Gundlach, H. (1993). *Entstehung und Gegenstand der Psychophysik*. Berlin: Springer.  
doi:10.1007/978-3-642-50262-0
- Hall, J. A. & Matsumoto, D. (2004). Gender differences in judgments of multiple emotions from facial expressions. *Emotion*, 4(2), 201–206. doi:10.1037/1528-3542.4.2.201

- Hancock, G. R. & Klockars, A. J. (1991). The effect of scale manipulations on validity: Targetting frequency rating scales for anticipated performance levels. *Applied Ergonomics*, 22(3), 147–154. doi:10.1016/0003-6870(91)90153-9
- Hand, D. J. (1996). Statistics and the theory of measurement. *Journal of the Royal Statistical Society*, 159(3), 445–492. doi:10.2307/2983326
- Harrell, F. E. (2015). *Regression modeling strategies: With applications to linear models, logistic and ordinal regression, and survival analysis* (2nd ed.). Cham: Springer. doi:10.1007/978-3-319-19425-7
- Harris, R. L. (1999). *Information graphics: A comprehensive illustrated reference*. New York, NY: Oxford University Press.
- Hartig, J., Frey, A. & Jude, N. (2012). Validität. In H. Moosbrugger & A. Kelava (Hrsg.), *Testtheorie und Fragebogenkonstruktion* (2. Aufl., S. 143–171). Berlin: Springer. doi:10.1007/978-3-642-20072-4\_7
- Harvey, A., Watkins, E., Mansell, W. & Shafran, R. (2004). *Cognitive behavioural processes across psychological disorders: A transdiagnostic approach to research and treatment*. Oxford: Oxford University Press.
- Harvey, N. & Bolger, F. (1996). Graphs versus tables: Effects of data presentation format on judgemental forecasting. *International Journal of Forecasting*, 12(1), 119–137. doi:10.1016/0169-2070(95)00634-6
- Harvey, R. J., Billings, R. S. & Nilan, K. J. (1985). Confirmatory factor analysis of the job diagnostic survey: Good news and bad news. *Journal of Applied Psychology*, 70(3), 461–468. doi:10.1037/0021-9010.70.3.461
- Heene, M., Hilbert, S., Draxler, C., Ziegler, M. & Bühner, M. (2011). Masking misfit in confirmatory factor analysis by increasing unique variances: A cautionary note on the usefulness of cutoff values of fit indices. *Psychological Methods*, 16(3), 319–336. doi:10.1037/a0024917
- Heymsfield, S. B., Scherzer, R., Pietrobelli, A., Lewis, C. E. & Grunfeld, C. (2009). Body mass index as a phenotypic expression of adiposity: Quantitative contribution of muscularity in a population-based sample. *International Journal of Obesity*, 33(12), 1363–1373. doi:10.1038/ijo.2009.184

- Heywood, H. B. (1931). On finite sequences of real numbers. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 134(824), 486–501. doi:10.1098/rspa.1931.0209
- Hilbert, S. (2016). *Eine empirische Untersuchung des Skalenniveaus von Fragebogenitems*. Beitrag präsentiert auf dem 50. Kongress der Deutschen Gesellschaft für Psychologie, 18.09.2016 – 22.09.2016, Leipzig.
- Hilbert, S., Küchenhoff, H., Sarubin, N., Nakagawa, T. T. & Bühner, M. (2016). The influence of the response format in a personality questionnaire: An analysis of a dichotomous, a Likert-type, and a visual analogue scale. *Testing, Psychometrics, Methodology in Applied Psychology*, 23(1), 3–24. doi:10.4473/TPM23.1.1
- Hilbert, S., Pargent, F., Naumann, F., Eichhorn, K., Ungar, P. & Bühner, M. (2018). *What's the measure? An empirical investigation of self-ratings on response scales*. Manuscript under review.
- Hinz, A., Brähler, E., Geyer, M. & Körner, A. (2003). Urteilseffekte beim NEO-FFI. *Diagnostica*, 49(4), 157–163. doi:10.1026//0012-1924.49.4.157
- Höfling, V., Moosbrugger, H., Schermelleh-Engel, K. & Heidenreich, T. (2011). Mindfulness or mindlessness?: A modified version of the Mindful Attention and Awareness Scale (MAAS). *European Journal of Psychological Assessment*, 27(1), 59–64. doi:10.1027/1015-5759/a000045
- Holm, K. (1974). Theorie der Frage. *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, 26(1), 91–114.
- Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*, 30(2), 179–185. doi:10.1007/BF02289447
- Hu, L. & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6(1), 1–55. doi:10.1080/10705519909540118
- Idaszak, J. R. & Drasgow, F. (1987). A revision of the Job Diagnostic Survey: Elimination of a measurement artifact. *Journal of Applied Psychology*, 72(1), 69–74. doi:10.1037/0021-9010.72.1.69

- Jackson, L. A., Sullivan, L. A. & Rostker, R. (1988). Gender, gender role, and body image. *Sex Roles*, 19(7/8), 429–443. doi:10.1007/BF00289717
- Jacob, R., Heinz, A. & Décieux, J. P. (2013). *Umfrage: Einführung in die Methoden der Umfrageforschung* (3. Aufl.). München: Oldenbourg.
- James, G., Witten, D., Hastie, T. & Tibshirani, R. (2013). *An introduction to statistical learning: With applications in R*. New York, NY: Springer New York. doi:10.1007/978-1-4614-7138-7
- Jeffery, R. W. & French, S. A. (1996). Socioeconomic status and weight control practices among 20- to 45-year-old women. *American Journal of Public Health*, 86(7), 1005–1010. doi:10.2105/AJPH.86.7.1005
- Jobe, J. & Loftus, E. (1991). Cognitive aspects of survey methodology. *Special issue of Applied Cognitive Psychology*, 5(3), 173–296.
- Jonkisz, E., Moosbrugger, H. & Brandt, H. (2012). Planung und Entwicklung von Tests und Fragebogen. In H. Moosbrugger & A. Kelava (Hrsg.), *Testtheorie und Fragebogenkonstruktion* (2. Aufl., S. 27–74). Berlin: Springer. doi:10.1007/978-3-642-20072-4\_3
- Kallus, K. W. (2016). *Erstellung von Fragebogen* (2. Aufl.). Wien: Facultas.
- Kastenbaum, R., Derbin, V., Sabatini, P. & Artt, S. (1972). "The ages of me": Toward personal and interpersonal definitions of functional aging. *The International Journal of Aging and Human Development*, 3(2), 197–211. doi:10.2190/TUJR-WTXK-866Q-8QU7
- Kelley, K. (2018). MBESS: The MBESS R package. R package (Version 4.4.3.) [Computersoftware]. Verfügbar unter <https://CRAN.R-project.org/package=MBESS>
- Kelley, T. L. (1927). *The interpretation of educational measurements*. New York, NY: World Book.
- Kersting, M. (2008). *Qualität in der Diagnostik und Personalauswahl – der DIN-Ansatz*. Göttingen: Hogrefe.
- Kersting, M., Häcker, H. O. & Hornke, L. F. (2011). Qualitätsstandards in der psychologischen Diagnostik. In L. F. Hornke, M. Amelang & M. Kersting (Hrsg.), *Grundfragen und Anwendungsfelder psychologischer Diagnostik* (S. 1–86). Göttingen: Hogrefe.

- Klatzky, R. L. (1998). Allocentric and egocentric spatial representations: Definitions, distinctions, and interconnections. In C. Freksa, C. Habel & K. F. Wender (Eds.), *Spatial cognition: An interdisciplinary approach to representation and processing of spatial knowledge* (pp. 1–17). Berlin: Springer. doi:10.1007/3-540-69342-4
- Kline, R. B. (2011). *Principles and practice of structural equation modeling* (3rd ed.). New York, NY: Guilford Press.
- Klein, R. A., Ratliff, K. A., Vianello, M., Adams, R. B. Jr., Bahník, Š., Bernstein, M. J., . . . Brumbaugh, C. C. (2014a). Data from investigating variation in replicability: A “many labs” replication project. *Journal of Open Psychology Data*, 2(1). e4. doi:10.5334/jopd.ad
- Klein, R. A., Ratliff, K. A., Vianello, M., Adams, R. B., Jr., Bahník, Š., Bernstein, M. J., . . . Nosek, B. A. (2014b). Investigating variation in replicability: A “many labs” replication project. *Social Psychology*, 45(3), 142–152. doi:10.1027/1864-9335/a000178
- Klein, R. A., Vianello, M., Hasselman, F., Adams, B. A., Adams, R. B. & Alper, S. (2015). Many labs 2: Investigating variation in replicability across sample and setting. Charlottesville, VA: Center for Open Science. Zugriff am 18.03.2018. Verfügbar unter <https://osf.io/8cd4r/>
- Kleinginna, P. R. & Kleinginna, A. M. (1981). A categorized list of emotion definitions, with suggestions for a consensual definition. *Motivation and Emotion*, 5(4), 345–379. doi:10.1007/BF00992553
- Klemm, O. (1911). *Geschichte der Psychologie*. Leipzig: Teubner.
- Knapp, T. R. (1990). Treating ordinal scales as interval scales: An attempt to resolve the controversy. *Nursing Research*, 39(2), 121–123. doi:10.1097/00006199-199003000-00019
- König, D. (2012). *Deutsche Version der Skala Rumination aus dem Rumination-Reflection Questionnaire (RRQ)*. Zugriff am 22.02.2018. Verfügbar unter [http://dk.akis.at/RRQ\\_Rumination.pdf](http://dk.akis.at/RRQ_Rumination.pdf)
- Konrath, S., Meier, B. P. & Bushman, B. J. (2014). Development and validation of the Single Item Narcissism Scale (SINS). *PLoS ONE*, 9(8), e103469, 1–15. doi:10.1371/journal.pone.0103469

- Krakauer, N. Y. & Krakauer, J. C. (2012). A new body shape index predicts mortality hazard independently of body mass index. *PLoS ONE*, 7(7), e39504.  
doi:10.1371/journal.pone.0039504
- Krantz, D. H., Luce, R. D., Suppes, P. & Tversky, A. (1971). *Foundations of measurement: Vol. I: Additive and polynomial representations*. New York, NY: Academic Press
- Krebs, D. & Matschinger, H. (1993). *Richtungseffekte von Itemformulierungen* (ZUMA-Arbeitsbericht, 1993/15). Mannheim: Zentrum für Umfragen, Methoden und Analysen -ZUMA-. Verfügbar unter <http://nbn-resolving.de/urn:nbn:de:0168-ssaar-70051>
- Krohne, H. W., Egloff, B., Kohlmann, C.-W. & Tausch, A. (1996). Untersuchungen mit einer deutschen Version der „Positive and Negative Affect Schedule“ (PANAS). *Diagnostica*, 42(2), 139–156.
- Krosnick, J. A. (1991). Response strategies for coping with the cognitive demands of attitude measures in surveys. *Applied Cognitive Psychology*, 5(3), 213–236.  
doi:10.1002/acp.2350050305
- Krosnick, J. A. (1999). Survey research. *Annual Review of Psychology*, 50, 537–567.  
doi:10.1146/annurev.psych.50.1.537
- Krosnick, J. A. & Berent, M. K. (1990). *The impact of verbal labeling of response alternatives and branching on attitude measurement reliability in surveys*. Paper presented at the annual meeting of the American Association for Public Opinion Research, February 1990, Lancaster, PA.
- Krosnick, J. A. & Presser, S. (2010). Question and Questionnaire design. In P. V. Marsden & J. D. Wright (Eds.), *Handbook of survey research* (2nd ed., pp. 263–313). Bingley, UK: Emerald Group.
- Kuder, G. F. & Richardson, M. W. (1937). The theory of the estimation of test reliability. *Psychometrika*, 2(3), 151–160. doi:10.1007/BF02288391
- Kuzon, W. M., Urbanchek, M. G. & McCabe, S. (1996). The seven deadly sins of statistical analysis. *Annals of Plastic Surgery*, 37(3), 265–272. doi:10.1097/00000637-199609000-00006



- Labovitz, S. (1967). Some observations on measurement and statistics. *Social Forces*, 46(2), 151–160. doi:10.2307/2574595
- Lang, F. R. & Lüdtke, O. (2005). Der Big Five-Ansatz der Persönlichkeitsforschung: Instrumente und Vorgehen. In S. Schumann (Hrsg.), *Persönlichkeit: Eine vergessene Größe der empirischen Sozialforschung* (S. 29–39). Wiesbaden: VS.
- Langeheine, R., Pannekoek, J. & van de Pol, F. (1996). Bootstrapping goodness-of-fit measures in categorical data analysis. *Sociological Methods and Research*, 24(4), 492–516. doi:10.1177/0049124196024004004
- Larson, M. R. (2000). Social desirability and self-reported weight and height. *International Journal of Obesity*, 24(5), 663–665. doi:10.1038/sj.ijo.0801233
- Leiner, D. J. (2014). SoSci Survey (Version 2.6.00-i) [Computersoftware]. Verfügbar unter <https://www.soscisurvey.de>
- Lenzner, T., Neuert, C. & Otto, W. (2015). *Kognitives Pretesting. Mannheim, GESIS – Leibniz-Institut für Sozialwissenschaften (SDM Survey Guidelines)*. doi:10.15465/sdm-sg\_010
- Lienert, G. A. & Raatz, U. (1998). *Testaufbau und Testanalyse* (6. Aufl.). Weinheim: Psychologie Verlags Union.
- Likert, R. (1932). A technique for the measurement of attitudes. *Archives of Psychology*, 22, 44–60.
- Lord, F. M. (1953). On the statistical treatment of football numbers. *American Psychologist*, 8(12), 750–751. doi:10.1037/h0063675
- Lozano, L. M., García-Cueto, E. & Muñiz, J. (2008). Effect of the number of response categories on the reliability and validity of rating scales. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences*, 4(2), 73–79. doi:10.1027/1614-2241.4.2.73
- MacCallum, R. C., Browne, M. W. & Sugawara, H. M. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods*, 1(2), 130–149. doi:10.1037/1082-989X.1.2.130
- Mandel, N. (2003). Shifting selves and decision making: The effects of self-construal priming on consumer risk-taking. *Journal of Consumer Research*, 30(1), 30–40. doi:10.1086/374700

- Manstead, A. S. R. & Livingstone, A. G. (2014). Forschungsmethoden in der Sozialpsychologie. In K. Jonas, W. Stroebe & M. Hewstone (Hrsg.), *Sozialpsychologie* (6. Aufl., S. 29–64). Berlin: Springer. doi:10.1007/978-3-642-41091-8\_2
- Marsh, H. W. (1984). Relations among dimensions of self-attribution, dimensions of self-concept, and academic achievements. *Journal of Educational Psychology*, 76(6), 1291–1308. doi:10.1037/0022-0663.76.6.1291
- Marsh, H. W. (1987). The big-fish-little-pond effect on academic self-concept. *Journal of Educational Psychology*, 79(3), 280–295. doi:10.1037/0022-0663.79.3.280
- Marsh, H. W. (1996). Positive and negative global self-esteem: A substantively meaningful distinction or artifactors? *Journal of Personality and Social Psychology*, 70(4), 810–819. doi:10.1037/0022-3514.70.4.810
- Marsh, H. W., Balla, J. R. & McDonald, R. P. (1988). Goodness-of-fit indexes in confirmatory factor analysis: The effect of sample size. *Psychological Bulletin*, 103(3), 391–410. doi:10.1037/0033-2909.103.3.391
- Marsh, H. W., Hau, K.-T., Balla, J. R. & Grayson, D. (1998). Is more ever too much? The number of indicators per factor in confirmatory factor analysis. *Multivariate Behavioral Research*, 33(2), 181–220. doi:10.1207/s15327906mbr3302\_1
- Marsh, H. W., Hau, K.-T. & Wen, Z. (2004). In search of golden rules: Comment on hypothesis-testing approaches to setting cutoff values for fit indexes and dangers in overgeneralizing Hu and Bentler's (1999) findings. *Structural Equation Modeling: A Multidisciplinary Journal*, 11(3), 320–341. doi:10.1207/s15328007sem1103\_2
- Marsh, H. W. & Parducci, A. (1978). Natural anchoring at the neutral point of category rating scales. *Journal of Experimental Social Psychology*, 14(2), 193–204. doi:10.1016/0022-1031(78)90025-2
- Marsh, H. W. & Parker, J. W. (1984). Determinants of student self-concept: Is it better to be a relatively large fish in a small pond even if you don't learn to swim as well? *Journal of Personality and Social Psychology*, 47(1), 213–231. doi:10.1037/0022-3514.47.1.213
- Martin, R. (1929). *Anthropometrie: Anleitung zu selbständigen anthropologischen Erhebungen* (2. Aufl.). Berlin: Springer.

- Matell, M. S. & Jacoby, J. (1971). Is there an optimal number of alternatives for Likert scale items? Study I: Reliability and validity. *Educational and Psychological Measurement*, 31(3), 657–674. doi:10.1177/001316447103100307
- Maydeu-Olivares, A., Kramp, U., García-Forero, C., Gallardo-Pujol, D. & Coffman, D. (2009). The effect of varying the number of response alternatives in rating scales: Experimental evidence from intra-individual effects. *Behavior Research Methods*, 41(2), 295–308. doi:10.3758/BRM.41.2.295
- McCrae, R. R., Costa, P. T. & Martin, T. A. (2005). The NEO-PI-3: A more readable revised NEO personality inventory. *Journal of Personality Assessment*, 84(3), 261–270. doi:10.1207/s15327752jpa8403\_05
- McDonald, R. P. (1970). The theoretical foundations of principal factor analysis, canonical factor analysis, and alpha factor analysis. *British Journal of Mathematical and Statistical Psychology*, 23, 1–21. doi:10.1111/j.2044-8317.1970.tb00432.x
- McPherson, J. & Mohr, P. (2005). The role of item extremity in the emergence of keying-related factors: An exploration with the Life Orientation Test. *Psychological Methods*, 10(1), 120–131. doi:10.1037/1082-989X.10.1.120
- Meisenberg, G. & Williams, A. (2008). Are acquiescent and extreme response styles related to low intelligence and education? *Personality and Individual Differences*, 44(7), 1539–1550. doi:10.1016/j.paid.2008.01.010
- Meloni, F. & Gana, K. (2001). Wording effects in the Italian version of the Penn State Worry Questionnaire. *Clinical Psychology & Psychotherapy*, 8(4), 282–287. doi:10.1002/cpp.294
- Mentzos, S. (2017). *Neurotische Konfliktverarbeitung: Einführung in die psychoanalytische Neurosenlehre unter Berücksichtigung neuer Perspektiven* (24. Aufl.). Frankfurt am Main: Fischer.
- Michalos, A. C. & Kahlke, P. M. (2010). Stability and sensitivity in perceived quality of life measures: Some panel results. *Social Indicators Research*, 98(3), 403–434. doi:10.1007/s11205-009-9554-2
- Miller, L. A. & Lovler, R. L. (2016). *Foundations of psychological testing: A practical approach* (5th ed.). Thousand Oaks, CA: Sage.

- Miller, M. B. (1995). Coefficient alpha: A basic introduction from the perspectives of classical test theory and structural equation modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, 2(3), 255–273. doi:10.1080/10705519509540013
- Miller, T. R. & Cleary, T. A. (1993). Direction of wording effects in balanced scales. *Educational and Psychological Measurement*, 53, 51–60. doi:10.1177/0013164493053001004
- Mitchell, A. J. & Coyne, J. C. (2007). Do ultra-short screening instruments accurately detect depression in primary care? A pooled analysis and meta-analysis of 22 studies. *British Journal of General Practice*, 57(535), 144–151.
- Molina, J. G., Rodrigo, M. F., Losilla, J.-M. & Vives, J. (2014). Wording effects and the factor structure of the 12-item General Health Questionnaire (GHQ-12). *Psychological Assessment*, 26(3), 1031–1037. doi:10.1037/a0036472
- Montepare, J. M. (1996a). Actual and subjective age-related differences in women's attitudes toward their bodies across the life span. *Journal of Adult Development*, 3(3), 171–182. doi:10.1007/BF02285777
- Montepare, J. M. (1996b). An assessment of adults' perceptions of their psychological, physical, and social ages. *Journal of Clinical Geropsychology*, 2(2), 117–128.
- Montepare, J. M. & Lachman, M. E. (1989). "You're only as old as you feel": Self-perceptions of age, fears of aging, and life satisfaction from adolescence to old age. *Psychology and Aging*, 4(1), 73–78. doi:10.1037/0882-7974.4.1.73
- Moosbrugger, H. & Kelava, A. (2012). Einführung und zusammenfassender Überblick. In H. Moosbrugger & A. Kelava (Hrsg.), *Testtheorie und Fragebogenkonstruktion* (2. Aufl., S. 1–4). Berlin: Springer. doi:10.1007/978-3-642-20072-4\_1
- Morf, C. C. & Koole, S. L. (2014). Das Selbst. In K. Jonas, W. Stroebe & M. Hewstone (Hrsg.), *Sozialpsychologie* (6. Aufl., S. 141–195). Berlin: Springer. doi:10.1007/978-3-642-41091-8\_5
- Mummendey, H. D. (2003). *Die Fragebogenmethode: Grundlagen und Anwendungen in Persönlichkeits-, Einstellungs- und Selbstkonzeptforschung* (4. Aufl.). Göttingen: Hogrefe.

- Mummendey, H. D. & Grau, I. (2014). *Die Fragebogen-Methode: Grundlagen und Anwendung in Persönlichkeits-, Einstellungs- und Selbstkonzeptforschung* (6. Aufl.). Göttingen: Hogrefe.
- Murphy, K. R. & Davidshofer, C. O. (2001). *Psychological testing: Principles and applications* (5th ed.). Upper Saddle River, NJ: Prentice Hall.
- Muthén, L. K. & Muthén, B. O. (1998–2017). *Mplus User's Guide*. (8th ed.). Los Angeles, CA: Muthén & Muthén
- Narens, L. (2007). *Introduction to the theories of measurement and meaningfulness and the use of symmetry in science*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Nevill, A. M., Stewart, A. D., Olds, T. & Holder, R. (2006). Relationship between adiposity and body size reveals limitations of BMI. *American Journal of Physical Anthropology*, 129(1), 151–156. doi:10.1002/ajpa.20262
- Neyer, F. J. & Asendorpf, J. B. (2018). *Psychologie der Persönlichkeit* (6. Aufl.). Berlin: Springer. doi:10.1007/978-3-662-54942-1
- Niederée, R. (1994). There is more to measurement than just measurement: Measurement theory, symmetry, and substantive theorizing. Review of foundations of measurement. Vol. 3: Representation, axiomatization, and invariance, by R. Duncan Luce, David H. Krantz, Patrick Suppes, and Amos Tversky. *Journal of Mathematical Psychology*, 38(4), 527–594. doi:10.1006/jmps.1994.1039
- Noelle, E. (1963). *Umfragen in der Massengesellschaft*. Reinbek: Rowohlt.
- Norman, G. (2010). Likert scales, levels of measurement and the "laws" of statistics. *Advances in Health Sciences Education: Theory and Practice*, 15(5), 625–632. doi:10.1007/s10459-010-9222-y
- Norušis, M. J. (2012). *IBM SPSS Statistics 19 advanced statistical procedures companion*. Upper Saddle River, NJ: Prentice Hall.
- Nunnally, J. C. (1978). *Psychometric theory* (2nd ed.). New York, NY: McGraw-Hill.
- Nunnally, J. C. & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). New York, NY: McGraw-Hill.
- Oaster, T. R. F. (1989). Number of alternatives per choice point and stability of Likert-type scales. *Perceptual and Motor Skills*, 68(2), 549–550. doi:10.2466/pms.1989.68.2.549

- Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, 349(6251), 1–8. doi:10.1126/science.aac4716
- Orth, B. (1974). *Einführung in die Theorie des Messens*. Stuttgart: Kohlhammer.
- Ostendorf, F. & Angleitner, A. (2004). *NEO-PI-R. NEO-Persönlichkeitsinventar nach Costa und McCrae: Revidierte Fassung*. Göttingen: Hogrefe.
- Pargent, F. (2017). *Detection, avoidance, and compensation – three studies on extreme response style* (Dissertation). Ludwig-Maximilians-Universität, München.  
urn:nbn:de:bvb:19-211562
- Pargent, F., Hilbert, S., Eichhorn, K. & Bühner, M. (2018). Can't make it better nor worse. An empirical study about the effectiveness of general rules of item construction on psychometric properties. *European Journal of Psychological Assessment*. Advance online publication. doi:10.1027/1015-5759/a000471
- Paulhus, D. L. (1984). Two-component models of socially desirable responding. *Journal of Personality and Social Psychology*, 46(3), 598–609. doi:10.1037/0022-3514.46.3.598
- Paulhus, D. L. (1991). Measurement and control of response bias. In J. P. Robinson, P. R. Shaver & L. S. Wrightsman (Eds.), *Measures of personality and social psychological attitudes: Volume 1 of measures of social psychological attitudes* (pp. 17–59). San Diego, CA: Academic Press.
- Payne, S. L. (1951). *The art of asking questions*. Princeton, NJ: Princeton University Press.
- Pilotte, W. J. & Gable, R. K. (1990). The impact of positive and negative item stems on the validity of a computer anxiety scale. *Educational and Psychological Measurement*, 50(3), 603–610. doi:10.1177/0013164490503016
- Playfair, W. (1786). *The commercial and political atlas. Representing, by means of stained copper-plate charts, the exports, imports, and general trade of England, at a single view. To which are added, charts of the revenue and debts of Ireland, done in the same manner by James Corry*. London: Debrett; Robinson; and Sewell.
- Plieninger, H. (2017). Mountain or molehill?: A simulation study on the impact of response styles. *Educational and Psychological Measurement*, 77(1), 32–53.  
doi:10.1177/0013164416636655
- Pongratz, L. J. (1984). *Problemgeschichte der Psychologie* (2. Aufl.) München: Francke.

- Porst, R. (2014). *Fragebogen: Ein Arbeitsbuch* (4. Aufl.). Wiesbaden: VS. doi:10.1007/978-3-658-02118-4
- Pospeschill, M. (2010). *Testtheorie, Testkonstruktion, Testevaluation*. München: Reinhardt.
- Preston, C. C. & Colman, A. M. (2000). Optimal number of response categories in rating scales: Reliability, validity, discriminating power, and respondent preferences. *Acta Psychologica*, 104(1), 1–15. doi:10.1016/S0001-6918(99)00050-5
- Prüfer, P. & Rexroth, M. (1996). Verfahren zur Evaluation von Survey-Fragen: Ein Überblick. *ZUMA-Nachrichten*, 39, 95–116.
- Prüfer, P. & Rexroth, M. (2005). Kognitive Interviews. *ZUMA How-to-Reihe*, Nr. 15.
- Raykov, T. (1997). Scale reliability, Cronbach's coefficient alpha, and violations of essential tau-equivalence with fixed congeneric components. *Multivariate Behavioral Research*, 32(4), 329–353. doi:10.1207/s15327906mbr3204\_2
- R Core Team (2016). foreign: Read data stored by Minitab, S, SAS, SPSS, Stata, Systat, Weka, dBase, .... R package (Version 0.8-67) [Computersoftware]. Verfügbar unter <https://CRAN.R-project.org/package=foreign>
- R Core Team (2017). R: A language and environment for statistical computing [Computersoftware]. *R Foundation for Statistical Computing, Vienna, Austria*. Verfügbar unter <https://www.R-project.org/>
- Revelle, W. (2017) psych: Procedures for personality and psychological research (Version 1.7.8.) [Computersoftware]. Northwestern University, Evanston, Illinois, USA. Verfügbar unter <https://CRAN.R-project.org/package=psych>
- Rhew, I. C., Simpson, K., Tracy, M., Lymp, J., McCauley, E., Tsuang, D. & Stoep, A. V. (2010). Criterion validity of the Short Mood and Feelings Questionnaire and one- and two-item depression screens in young adolescents. *Child and Adolescent Psychiatry and Mental Health*, 4(8), 1–11. doi:10.1186/1753-2000-4-8
- Richter, M. & Schaefer, K. (2009). *Der deutsche Fußreport 2009. Auswertung und Ergebnisse der jüngsten, bundesweiten Fuß- und Beinmessaktion* [Datendatei].
- Rizopoulos, D. (2006). ltm: An R package for latent variable modelling and item response theory analyses. *Journal of Statistical Software*, 17(5), 1–25 [Computersoftware]. Verfügbar unter URL <http://www.jstatsoft.org/v17/i05/>

- Rohrmann, B. (1978). Empirische Studien zur Entwicklung von Antwortskalen für die sozialwissenschaftliche Forschung. *Zeitschrift für Sozialpsychologie*, 9(3), 222–245.
- Rosch, E. H. (1973). On the internal structure of perceptual and semantic categories. In T. E. Moore (Ed.), *Cognitive development and the acquisition of language* (pp. 111–144). New York, NY: Academic Press.
- Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, 48(2), 1–36 [Computersoftware]. Verfügbar unter <http://www.jstatsoft.org/v48/i02/>.
- Rost, J. (1999). Was ist aus dem Rasch-Modell geworden? *Psychologische Rundschau*, 50(3), 140–156. doi:10.1026//0033-3042.50.3.140
- Rost, J. (2004). *Lehrbuch Testtheorie – Testkonstruktion* (2. Aufl.). Bern: Huber.
- Rowe, D. C., Jacobson, K. C. & van den Oord, E. J. C. G. (1999). Genetic and environmental influences on vocabulary IQ: Parental education level as moderator. *Child Development*, 70(5), 1151–1162. doi:10.1111/1467-8624.00084
- Rowland, M. L. (1990). Self-reported weight and height. *The American Journal of Clinical Nutrition*, 52(6), 1125–1133. doi:10.1093/ajcn/52.6.1125
- Ruff, C. B. (2000). Body mass prediction from skeletal frame size in elite athletes. *American Journal of Physical Anthropology*, 113(4), 507–517. doi:10.1002/1096-8644(200012)113:4<507::AID-AJPA5>3.0.CO;2-F
- Russell, S. J., Schifter, D. & Bastable, V. (2002). *Working with data: Statistics casebook. Developing mathematical ideas*. Parsippany, NJ: Dale Seymour.
- Saint-Mont, U. (2011). *Statistik im Forschungsprozess: Eine Philosophie der Statistik als Baustein einer integrativen Wissenschaftstheorie*. Heidelberg: Springer. doi:10.1007/978-3-7908-2723-1
- Sargent, J. D. & Blanchflower, D. G. (1994). Obesity and stature in adolescence and earnings in young adulthood: Analysis of a British birth cohort.. *Archives of Pediatrics & Adolescent Medicine*, 148(7), 681–687. doi:10.1001/archpedi.1994.02170070019003
- Sarlio-Lähteenkorva, S. & Lahelma, E. (1999). The association of body mass index with social and economic disadvantage in women and men. *International Journal of Epidemiology*, 28(3), 445–449. doi:10.1093/ije/28.3.445



- Scheuch, E. K. (1962). Das Interview in der Sozialforschung. In R. König (Hrsg.), *Handbuch der empirischen Sozialforschung. Band 2* (S. 66–190). Stuttgart: Enke.
- Schlegel, B. & Steenbergen, M. (2017). brant: Test for parallel regression assumption. R package (Version 0.1-3) [Computersoftware]. Verfügbar unter <https://CRAN.R-project.org/package=brant>
- Schmithüsen, F. & Krampen, G. (2015). Geschichte der Psychologie. In F. Schmithüsen (Hrsg.), *Lernskript Psychologie: Die Grundlagenfächer kompakt* (S. 1–19). Berlin: Springer. doi:10.1007/978-3-662-44941-7\_1
- Schmitt, N. & Stults, D. M. (1985). Factors defined by negatively keyed items: The result of careless respondents? *Applied Psychological Measurement*, 9(4), 367–373. doi:10.1177/014662168500900405
- Schmitt, N. & Stults, D. M. (1986). Methodology review: Analysis of multitrait–multimethod matrices. *Applied Psychological Measurement*, 10(1), 1–22. doi:10.1177/014662168601000101
- Schmitz, N., Hartkamp, N., Baldini, C., Rollnik, J. & Tress, W. (2001). Psychometric properties of the German version of the NEO-FFI in psychosomatic outpatients. *Personality and Individual Differences*, 31(5), 713–722. doi:10.1016/S0191-8869(00)00173-2
- Schneider, H. & Franke, B. (2014). *Bildungsentscheidungen von Studienberechtigten. Studienberechtigte 2012 ein halbes Jahr vor und ein halbes Jahr nach Schulabschluss* (Forum Hochschule 6|2014). Hannover: DZHW.
- Schnell, R., Hill, P. B. & Esser, E. (2013). *Methoden der empirischen Sozialforschung* (10. Aufl.). München: Oldenbourg.
- Schriesheim, C. A., Eisenbach, R. J. & Hill, K. D. (1991). The effect of negation and polar opposite item reversals on questionnaire reliability and validity: An experimental investigation. *Educational and Psychological Measurement*, 51(1), 67–78. doi:10.1177/0013164491511005
- Schriesheim, C. A. & Hill, K. D. (1981). Controlling acquiescence response bias by item reversals: The effect on questionnaire validity. *Educational and Psychological Measurement*, 41(4), 1101–1114. doi:10.1177/001316448104100420

- Schuldt, J. P., Roh, S. & Schwarz, N. (2015). Questionnaire design effects in climate change surveys: Implications for the partisan divide. *The ANNALS of the American Academy of Political and Social Science*, 658(1), 67–85. doi:10.1177/0002716214555066
- Schuman, H. & Presser, S. (1981). *Questions and answers in attitude surveys: Experiments on question form, wording, and context*. New York, NY: Academic Press.
- Schwall, A. R. (2012). Defining age and using age-relevant constructs. In J. W. Hedge & W. C. Borman (Eds.), *The Oxford handbook of work and aging* (pp. 169–186). New York, NY: Oxford University Press.
- Schwarz, N. (1999). Self-reports: How the questions shape the answers. *American Psychologist*, 54(2), 93–105. doi:10.1037/0003-066X.54.2.93
- Schwarz, N. (2007). Cognitive aspects of survey methodology. *Applied Cognitive Psychology*, 21(2), 277–287. doi:10.1002/acp.1340
- Schwarz, N. & Hippler, H.-J. (1995). The numeric values of rating scales: A comparison of their impact in mail surveys and telephone interviews. *International Journal of Public Opinion Research*, 7(1), 72–74. doi:10.1093/ijpor/7.1.72
- Schwarz, N., Knäuper, B., Hippler, H.-J., Noelle-Neumann, E. & Clark, L. (1991). Rating scales: Numeric values may change the meaning of scale labels. *Public Opinion Quarterly*, 55(4), 570–582. doi:10.1086/269282
- Shah, P. (2002). Graph comprehension: The role of format, content and individual differences. In M. Anderson, B. Meyer & P. Olivier (Eds.), *Diagrammatic representation and reasoning* (pp. 173–185). London: Springer. doi:10.1007/978-1-4471-0109-3\_10
- Shah, P. & Hoeffner, J. (2002). Review of graph comprehension research: Implications for instruction. *Educational Psychology Review*, 14(1), 47–69. doi:10.1023/A:1013180410169
- Shepperd, J. A. & Strathman, A. J. (1989). Attractiveness and height: The role of stature in dating preference, frequency of dating, and perceptions of attractiveness. *Personality and Social Psychology Bulletin*, 15(4), 617–627. doi:10.1177/0146167289154014

- Shih, M., Pittinsky, T. L. & Ambady, N. (1999). Stereotype susceptibility: Identity salience and shifts in quantitative performance. *Psychological Science*, 10(1), 80–83.  
doi:10.1111/1467-9280.00111
- Siegel, S. & Castellan, N. J. (1988). *Nonparametric statistics for the behavioral sciences* (2nd ed.). New York, NY: McGraw-Hill.
- Simon, B. (2012). A place in the world: Self and social categorization. In T. R. Tyler, R. M. Kramer & O. P. John (Eds.), *The psychology of the social self* (pp. 47–69). New York, NY: Psychology Press.
- Simon, H. A. (1959). Theories of decision-making in economics and behavioral science. *The American Economic Review*, 49(3), 253–283.
- Sočan, G. (2000). Assessment of reliability when test items are not essentially  $\tau$ -equivalent. *Developments in Survey Methodology*, 15, 23–35.
- Stevens, S. S. (1946). On the theory of scales of measurement. *Science*, 103(2684), 677–680.  
doi:10.1126/science.103.2684.677
- Steyer, R. & Eid, M. (1993). *Messen und Testen*. Berlin: Springer. doi:10.1007/978-3-642-97455-7
- Sturgis, P., Roberts, C. & Smith, P. (2014). Middle alternatives revisited: How the neither/nor response acts as a way of saying “I don’t know”? *Sociological Methods & Research*, 43(1), 15–38. doi:10.1177/0049124112452527
- Sudman, S., Bradburn, N. M. & Schwarz, N. (1996). *Thinking about answers: The application of cognitive processes to survey methodology*. San Francisco, CA: Jossey-Bass.
- Sudman, S. & Wansink, B. (2002). *Consumer panels* (2nd ed.). Chicago, IL: American Marketing Association.
- Swain, S. D., Weathers, D. & Niedrich, R. W. (2008). Assessing three sources of misresponse to reversed Likert items. *Journal of Marketing Research*, 45(1), 116–131.  
doi:10.1509/jmkr.45.1.116
- Swami, V. & Furnham, A. (2008). *The psychology of physical attraction*. London: Routledge.
- Tourangeau, R., Rips, L. J. & Rasinski, K. (2000). *The psychology of survey response*. Cambridge: Cambridge University Press.

- Triandis, H. C. (2001). Individualism-collectivism and personality. *Journal of Personality*, 69(6), 907–924. doi:10.1111/1467-6494.696169
- Triandis, H. C., Leung, K., Villareal, M. J. & Clack, F. L. (1985). Allocentric versus idiocentric tendencies: Convergent and discriminant validation. *Journal of Research in Personality*, 19(4), 395–415. doi:10.1016/0092-6566(85)90008-X
- Tukey, J. W. (1962). The future of data analysis. *The Annals of Mathematical Statistics*, 33(1), 1–67. doi:10.1214/aoms/1177704711
- Tutz, G. (2012). *Regression for categorical data. Cambridge series in statistical and probabilistic mathematics*. Cambridge: Cambridge University Press.
- Tzamaloukas, A. H., Patron, A. & Malhotra, D. (1994). Body mass index in amputees. *Journal of Parenteral and Enteral Nutrition*, 18(4), 355–358. doi:10.1177/014860719401800414
- Van der Linden, W. J. (2016a). Unidimensional Logistic Response Models. In W. J. van der Linden (Ed.), *Handbook of item response theory. Volume one: Models* (pp. 13–30). Boca Raton, FL: CRC Press.
- Van der Linden, W. J. (Ed.). (2016b). *Handbook of item response theory. Volume one: Models*. Boca Raton, FL: CRC Press.
- Van Vaerenbergh, Y. & Thomas, T. D. (2013). Response styles in survey research: A literature review of antecedents, consequences, and remedies. *International Journal of Public Opinion Research*, 25(2), 195–217. doi:10.1093/ijpor/eds021
- Velleman, P. F. & Wilkinson, L. (1993). Nominal, ordinal, interval, and ratio typologies are misleading. *The American Statistician*, 47(1), 65–72. doi:10.1080/00031305.1993.10475938
- Venables, W. N. & Ripley, B. D. (2002) Modern applied statistics with s (4th ed.) [Computersoftware]. New York, NY: Springer. Verfügbar unter <http://www.stats.ox.ac.uk/pub/MASS4>
- Wahlster, W. (1977). *Die Repräsentation von vagem Wissen in natürlichsprachlichen Systemen der Künstlichen Intelligenz*. Bericht Nr. 38, Hamburg: Institut für Informatik.

- Wang, L. & Lin, W. (2011). Wording effects and the dimensionality of the General Health Questionnaire (GHQ-12). *Personality and Individual Differences*, 50(7), 1056–1061. doi:10.1016/j.paid.2011.01.024
- Wang, W.-C., Chen, H.-F. & Jin, K.-Y. (2015). Item response theory models for wording effects in mixed-format scales. *Educational and Psychological Measurement*, 75(1), 157–178. doi:10.1177/0013164414528209
- Weber, W. & Wenzel, A. (2013). Interaktive Infografiken: Standortbestimmung und Definition. In W. Weber, M. Burmester & R. Tille (Hrsg.), *Interaktive Infografiken* (S. 3–23). Berlin: Springer. doi:10.1007/978-3-642-15453-9
- Weijters, B. & Baumgartner, H. (2012). Misresponse to reversed and negated items in surveys: A review. *Journal of Marketing Research*, 49(5), 737–747. doi:10.1509/jmr.11.0368
- Weijters, B., Cabooter, E. & Schillewaert, N. (2010). The effect of rating scale format on response styles: The number of response categories and response category labels. *International Journal of Research in Marketing*, 27(3), 236–247. doi:10.1016/j.ijresmar.2010.02.004
- Weijters, B., Geuens, M. & Schillewaert, N. (2010). The stability of individual response styles. *Psychological Methods*, 15(1), 96–110. doi:10.1037/a0018721
- Weng, L.-J. (2004). Impact of the number of response categories and anchor labels on coefficient alpha and test-retest reliability. *Educational and Psychological Measurement*, 64(6), 956–972. doi:10.1177/0013164404268674
- Weston, R. & Gore, P. A. (2006). A brief guide to structural equation modeling. *The Counseling Psychologist*, 34(5), 719–751. doi:10.1177/0011000006286345
- Wetzel, E. & Carstensen, C. H. (2017). Multidimensional modeling of traits and response styles. *European Journal of Psychological Assessment*, 33, 352–364. doi:10.1027/1015-5759/a000291
- Wetzel, E., Carstensen, C. H. & Böhnke, J. R. (2013). Consistency of extreme response style and non-extreme response style across traits. *Journal of Research in Personality*, 47(2), 178–189. doi:10.1016/j.jrp.2012.10.010

- Wetzel, E., Lüdtke, O., Zettler, I. & Böhnke, J. R. (2016). The stability of extreme response style and acquiescence over 8 years. *Assessment*, 23(3), 279–291.  
doi:10.1177/1073191115583714
- Wickham, H. (2009). ggplot2: Elegant graphics for data analysis [Computersoftware]. New York, NY: Springer. Verfügbar unter <http://ggplot2.org>
- Wickham, H. & Francois, R. (2016). dplyr: A grammar of data manipulation. R package (Version 0.5.0) [Computersoftware]. Verfügbar unter <https://CRAN.R-project.org/package=dplyr>
- Williams, J. W., Mulrow, C. D., Kroenke, K., Dhanda, R., Badgett, R. G., Omori, D. & Lee, S. (1999). Case-finding for depression in primary care: A randomized trial. *The American Journal of Medicine*, 106(1), 36–43. doi:10.1016/S0002-9343(98)00371-4
- Willis, G. B. (2005). *Cognitive interviewing: A tool for improving questionnaire design*. Thousand Oaks, CA: Sage.
- Ziebland, S., Thorogood, M., Fuller, A. & Muir, J. (1996). Desire for the body normal: Body image and discrepancies between self reported and measured height and weight in a British population. *Journal of Epidemiology & Community Health*, 50(1), 105–106.  
doi:10.1136/jech.50.1.105
- Zinbarg, R. E., Revelle, W., Yovel, I. & Li, W. (2005). Cronbach's  $\alpha$ , Revelle's  $\beta$ , and McDonald's  $\omega_H$ : Their relations with each other and two alternative conceptualizations of reliability. *Psychometrika*, 70(1), 123–133.  
doi:10.1007/s11336-003-0974-7

## Anhang

### A. Anhang zu Studie I

Tabelle A.1.

*Items zur Selbstbewertung auf fünf Merkmalsdimensionen*

Merkmal und Item-Nummer		Item
Alter	1	Ich bin alt.
	1(a)	Ich halte mich für alt.
	1(b)	Andere halten mich für alt.
Körpergröße	2	Ich bin groß.
	2(a)	Ich halte mich für groß.
	2(b)	Andere halten mich für groß.
Schuhgröße	3	Ich habe große Füße.
	3(a)	Ich halte meine Füße für groß.
	3(b)	Andere halten meine Füße für groß.
Körpergewicht	4	Ich wiege viel.
	4(a)	Ich finde, dass ich viel wiege.
	4(b)	Andere finden, dass ich viel wiege.
Einkommen	5	Ich verdiene viel.
	5(a)	Ich finde, dass ich viel verdiene.
	5(b)	Andere finden, dass ich viel verdiene.

**Anmerkungen.**

Die Nummerierung der Items entspricht der Nummerierung im Fragebogen.

Die Nummerierung der Items war für die Teilnehmer im Fragebogen nicht sichtbar.

Die Reihenfolge der Merkmale entspricht der Abfragereihenfolge im Fragebogen.

(a) Egozentrische Perspektive; (b) Allozentrische Perspektive; Tabelle entnommen aus Eichhorn (2016).

Abbildung A.1. Balkendiagramme für das Merkmal Alter

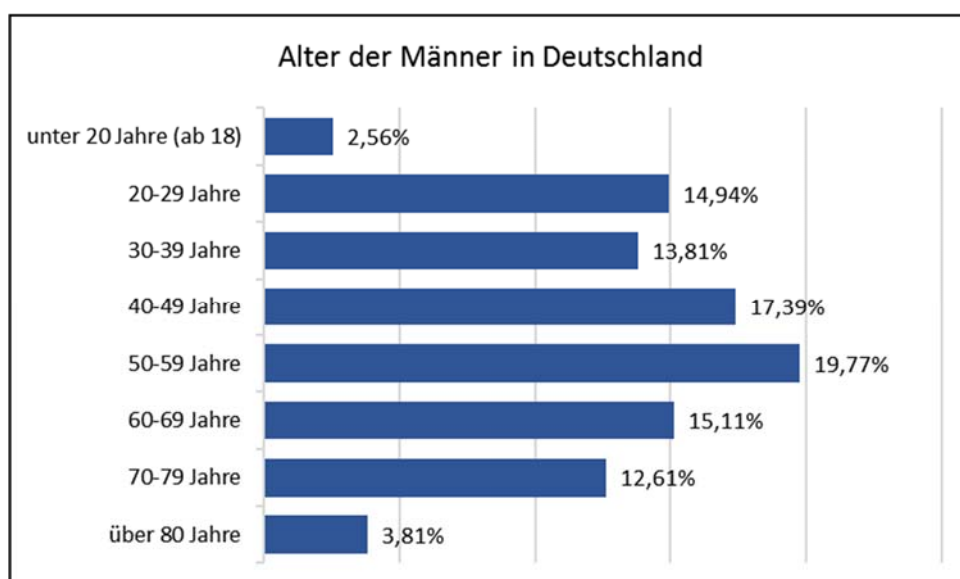
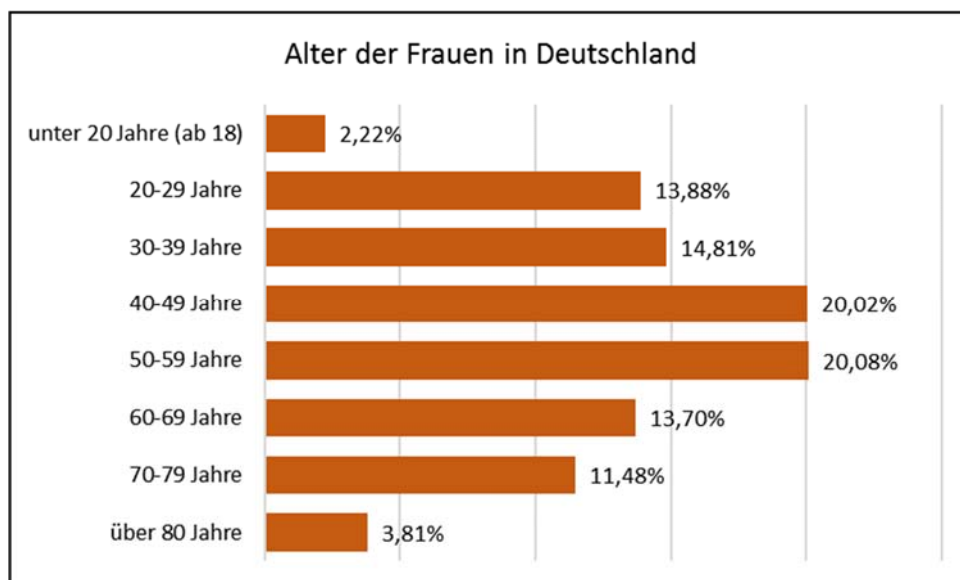
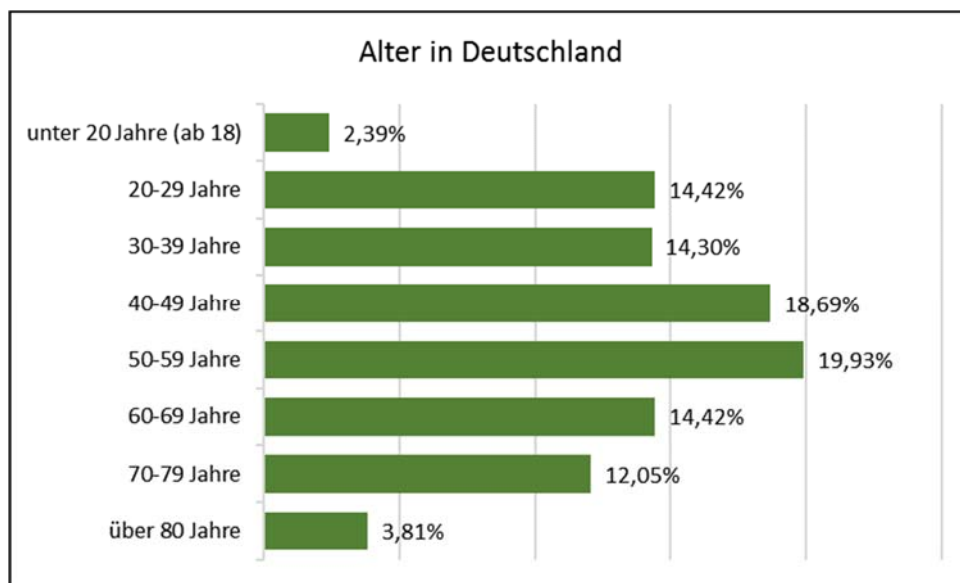




Abbildung A.2. Balkendiagramme für das Merkmal Körpergröße

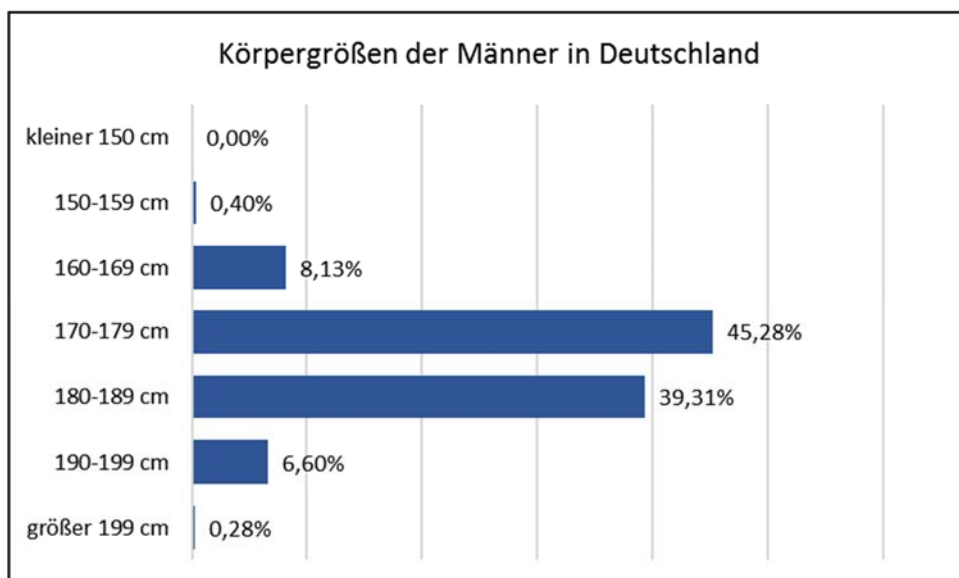
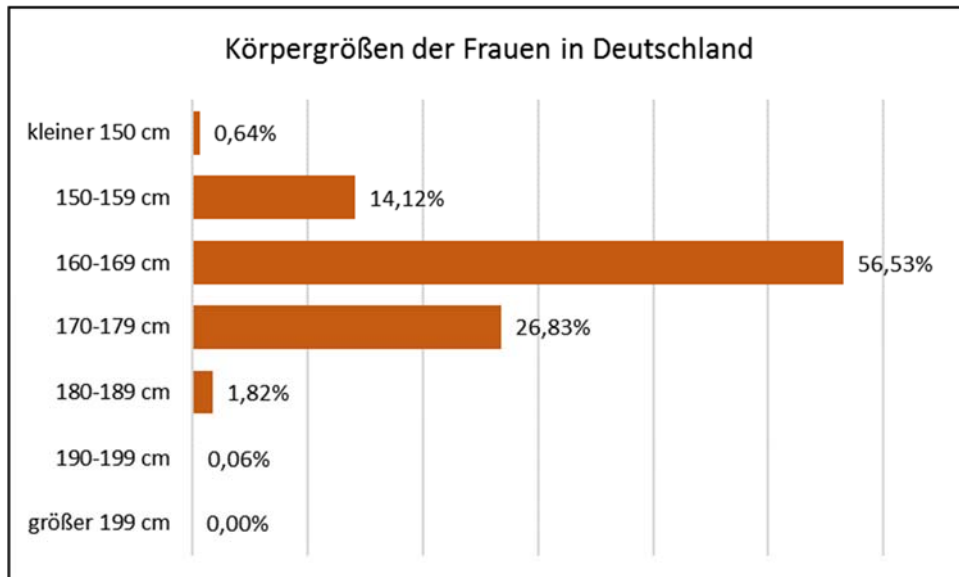
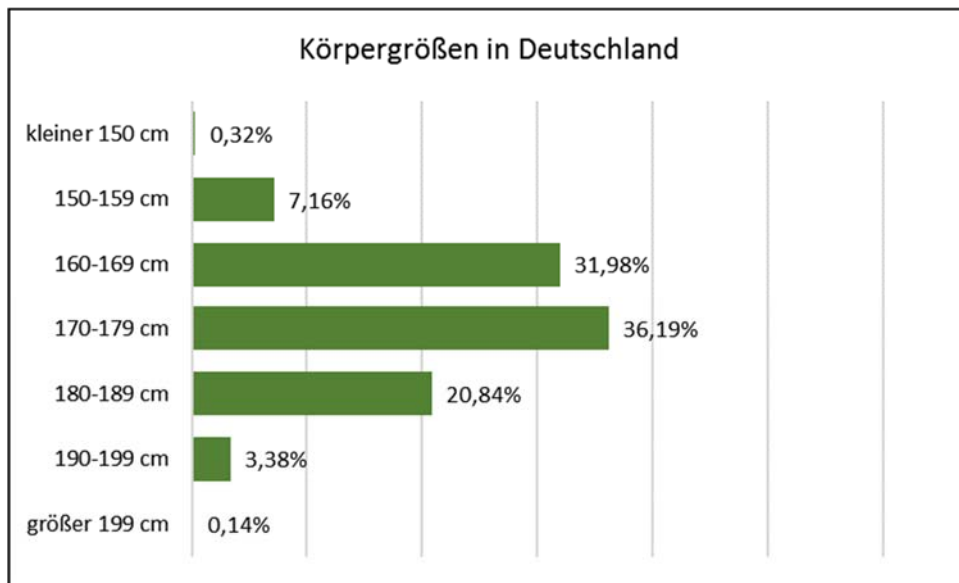


Abbildung A.3. Balkendiagramme für das Merkmal Schuhgröße

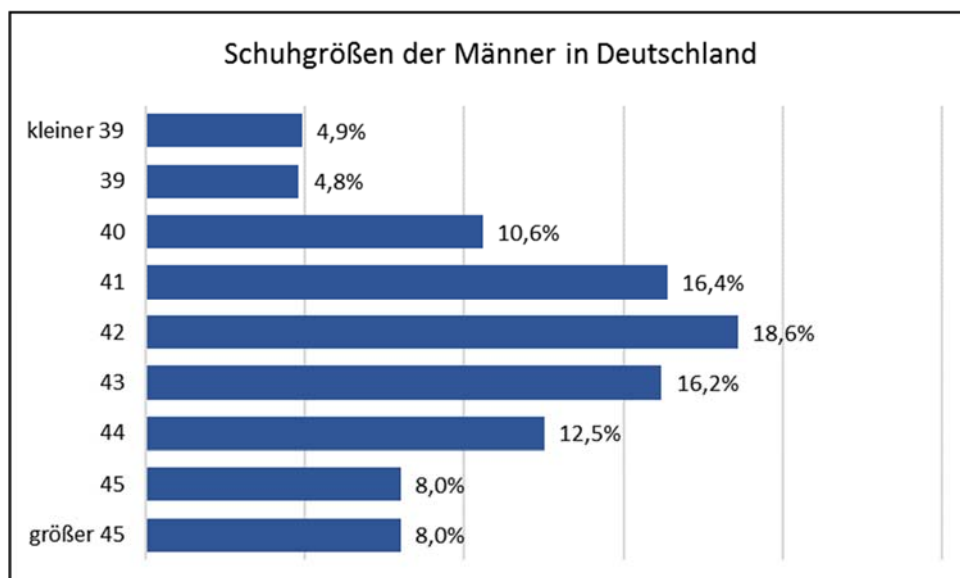
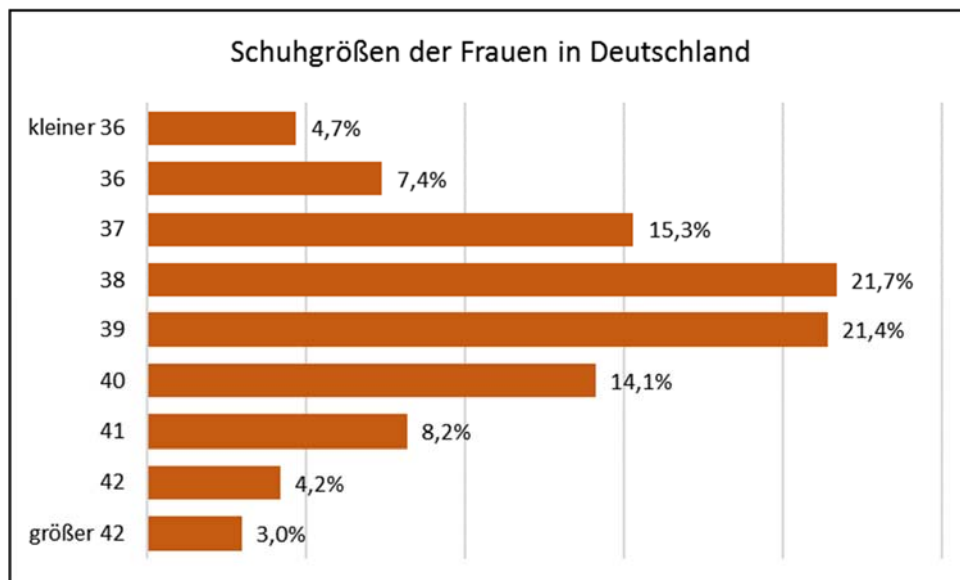
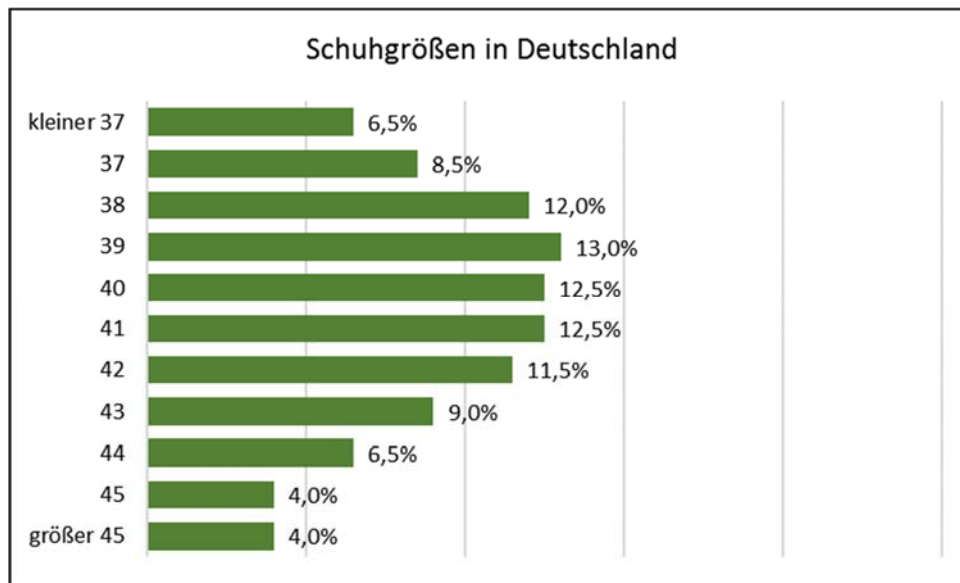


Abbildung A.4. Balkendiagramme für das Merkmal Körpergewicht

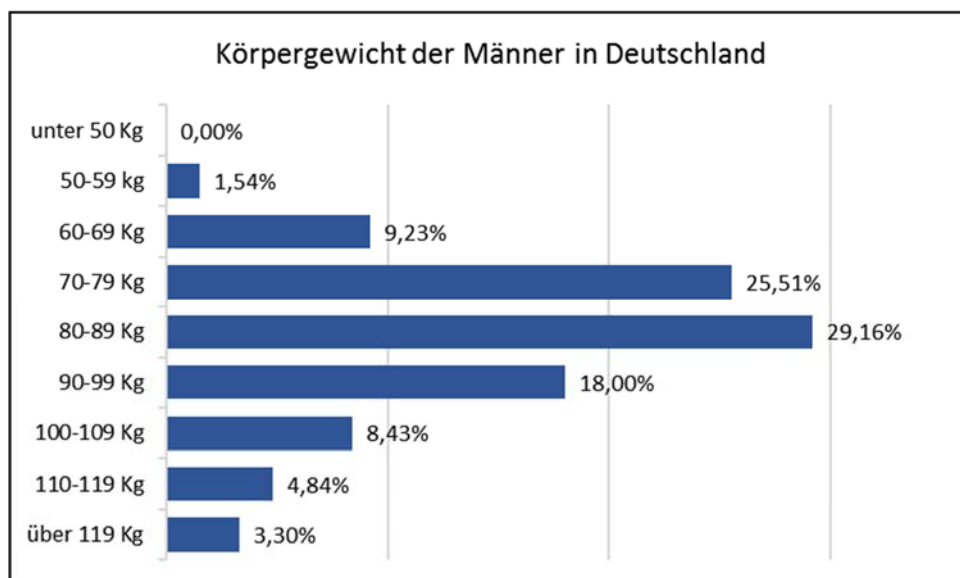
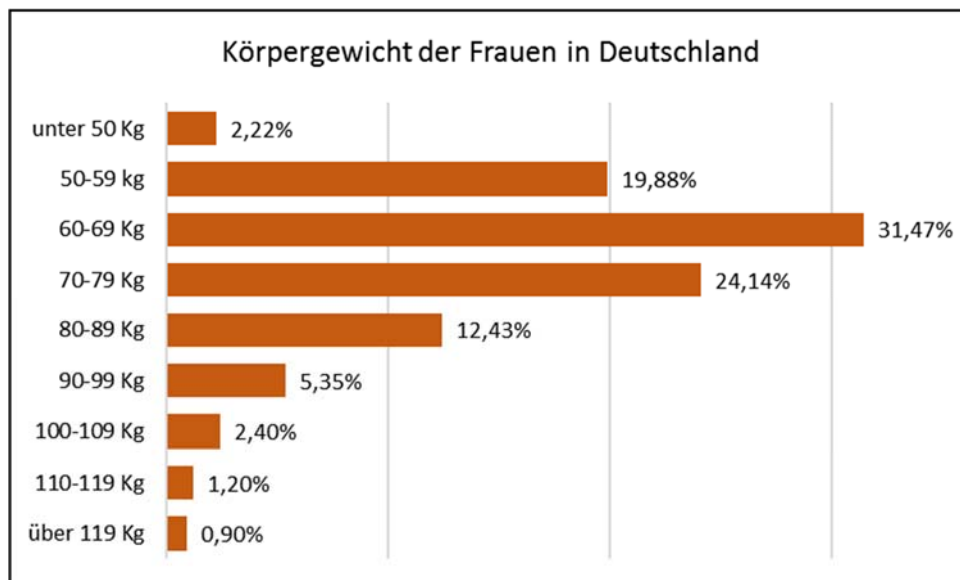
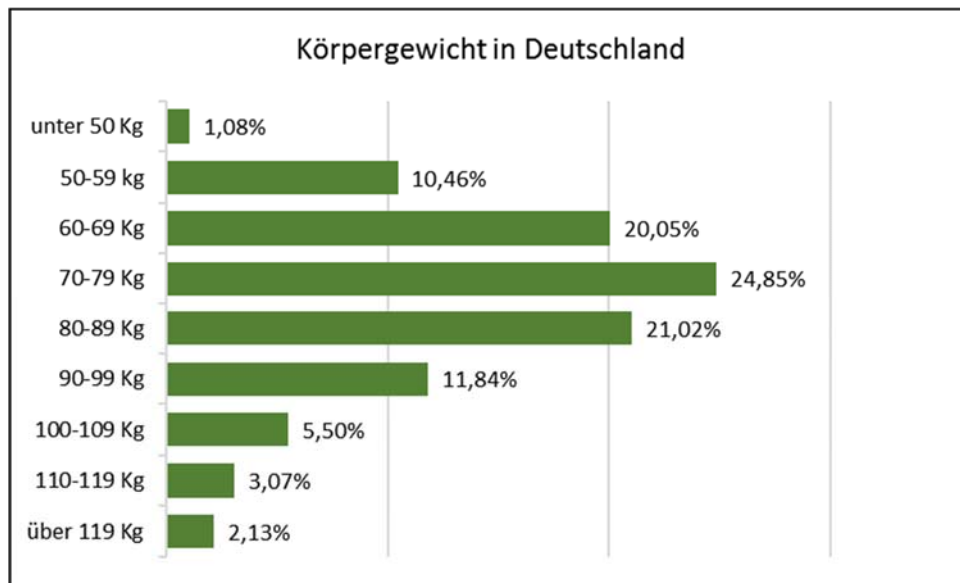


Abbildung A.5. Balkendiagramme für das Merkmal Einkommen

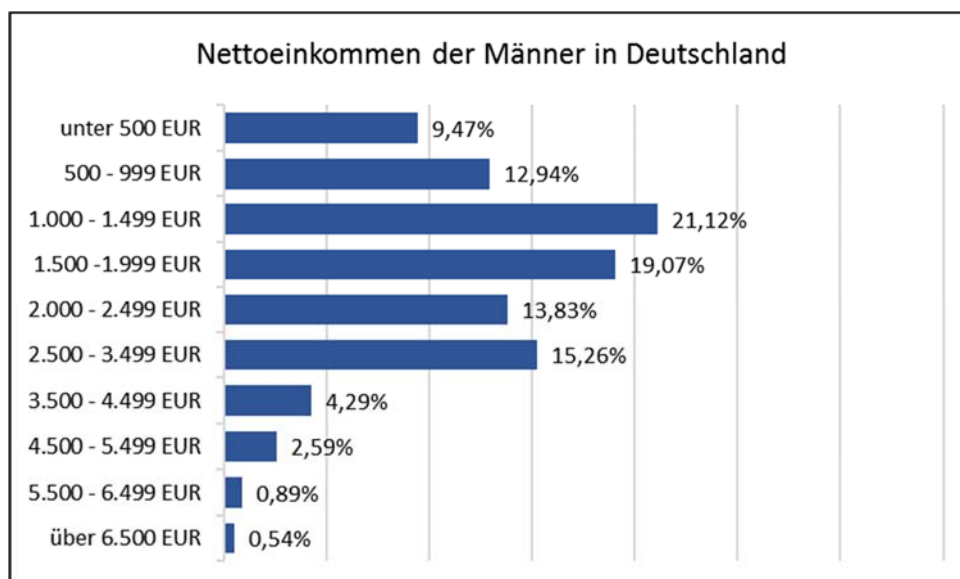
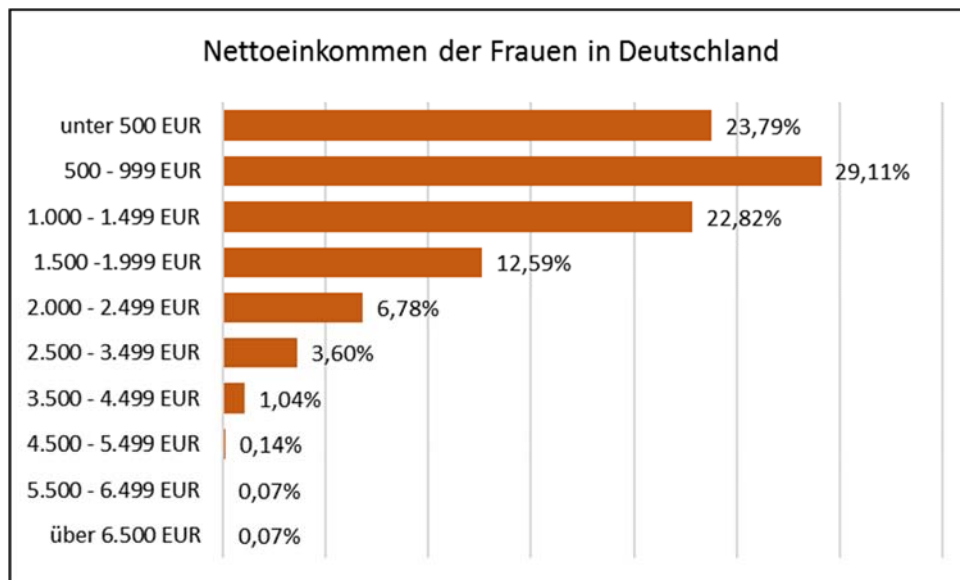
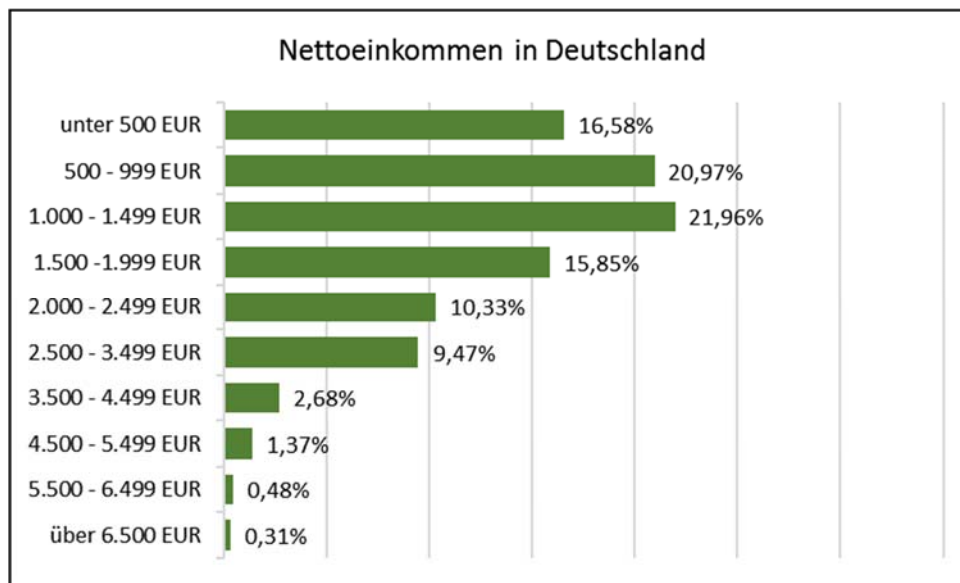


Abbildung A.6. Screenshot Onlinefragebogen der Version GV/GSV-Frauen

### Informationen zur Befragung

Sehr geehrte Teilnehmerin, sehr geehrter Teilnehmer,

bitte lesen Sie die folgenden Informationen zur Befragung und zum Schutz der erhobenen Daten durch.

Die Studie dauert ca. **3 Minuten**.

Ziel der Befragung ist die Evaluierung von psychologischen Messinstrumenten im Rahmen meiner Masterarbeit im Fach Psychologie an der Ludwig-Maximilians-Universität München.

#### 1. Daten

In dem Fragebogen werden Ihre Daten anonym behandelt, Ihr Name wird an keiner Stelle erfasst. Die aufgezeichneten Daten lassen sich in folgende Kategorien unterteilen:

1. Demografische Daten
2. Antworten auf Fragebogenitems

Sie werden zunächst um die Angabe **konkreter Werte** gebeten. Anschließend werden Sie nach Ihrer Zustimmung bzw. Ablehnung zu vorgegebenen Aussagen gefragt, die Sie auf einer Skala von 1 (= völlig **unzutreffend**) bis 5 (= völlig **zutreffend**) angeben sollen.

#### 2. Aufbewahrung und Löschung der Daten

Die anonymisierten Daten (Demographische Daten, Antworten auf Fragebogenitems) werden auf Datenträgern von der Versuchsleiterin und entsprechenden Lehrpersonen gespeichert und sind nicht öffentlich zugänglich. Sie verbleiben in den Händen der Versuchsleiterin bzw. den entsprechenden Lehrpersonen und werden nicht an Dritte weitergegeben. Die Verwendung erfolgt im Rahmen der Masterarbeit der Versuchsleiterin. Die VersuchsteilnehmerInnen können jederzeit und ohne Nennung von Gründen die Befragung abbrechen.

**Wenn Sie mit den oben genannten Punkten zur Datenerhebung einverstanden sind, geht es nun weiter zum Fragebogen. Sollten Sie mit den oben genannten Punkten nicht einverstanden sein, können Sie die Befragung an dieser Stelle beenden.**

Gerne können Sie Fragen an die Versuchsleiterin richten: [Kathryn.Eichhorn@campus.lmu.de](mailto:Kathryn.Eichhorn@campus.lmu.de)

Weiter

Abbildung A.7. Screenshot Onlinefragebogen der Version GV/GSV-Frauen

**1. Wie alt sind Sie?**

Ich bin  Jahre alt.

**2. Was ist Ihr höchster Bildungsabschluss?**

- ☐ Kein Schulabschluss
- ☐ Hauptschulabschluss
- ☐ Mittlere Reife
- ☐ Fachabitur oder Abitur
- ☐ Hochschulabschluss oder Universitätsabschluss

**3. Wie groß sind Sie (in cm)?**

Ich bin  cm groß.

**4. Welche Schuhgröße haben Sie (europäische Größenangabe)?**

Meine Schuhgröße ist .

**5. Wieviel wiegen Sie (in kg)?**

Ich wiege  kg.

**6. Wie hoch ist Ihr monatliches Nettoeinkommen (in Euro)?**

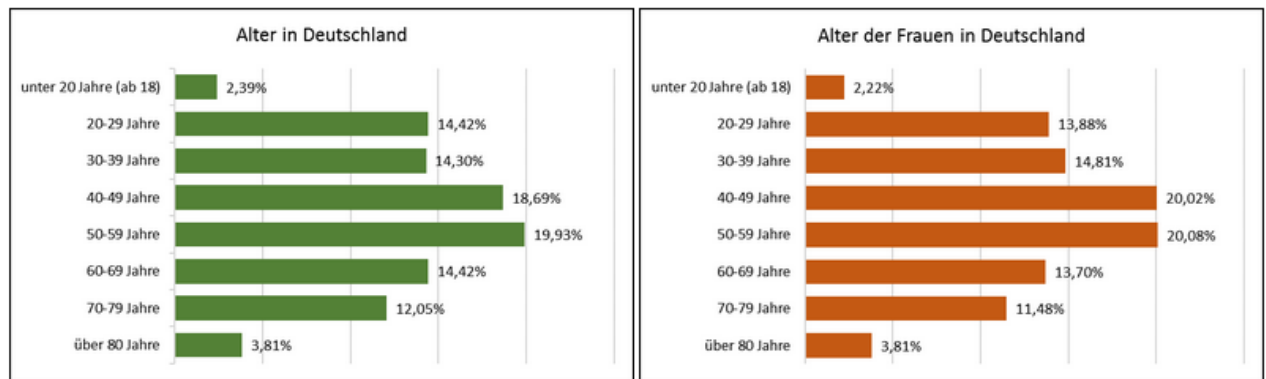
Mein monatliches Nettoeinkommen beträgt  EUR.

**7. Welches Geschlecht haben Sie?**

- ☐ weiblich
- ☐ männlich

Weiter

Abbildung A.8. Screenshot Onlinefragebogen der Version GV/GSV-Frauen



#### 8. Fragen zum Alter

Ich bin alt

völlig unzutreffend    unzutreffend    weder noch    zutreffend    völlig zutreffend

1                      2                      3                      4                      5

☐                      ☐                      ☐                      ☐                      ☐

Ich halte mich für alt.

völlig unzutreffend    unzutreffend    weder noch    zutreffend    völlig zutreffend

1                      2                      3                      4                      5

☐                      ☐                      ☐                      ☐                      ☐

Andere halten mich für alt.

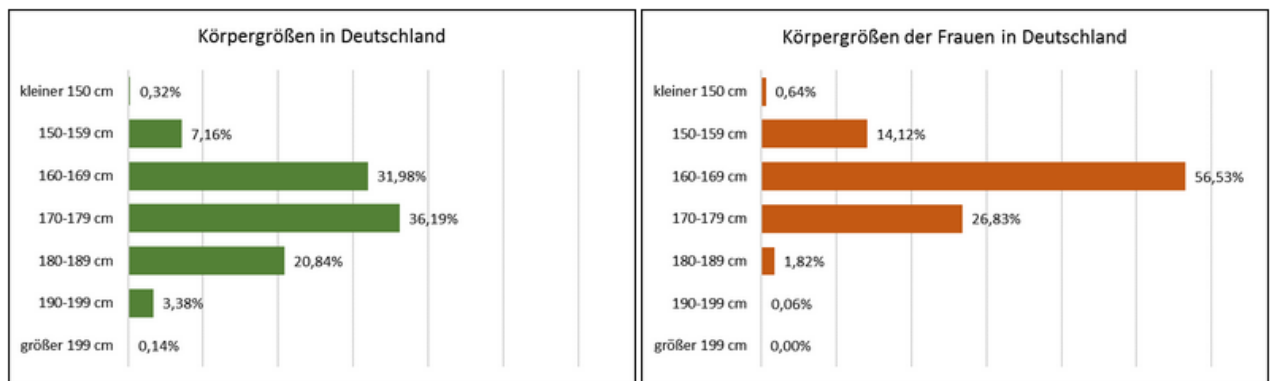
völlig unzutreffend    unzutreffend    weder noch    zutreffend    völlig zutreffend

1                      2                      3                      4                      5

☐                      ☐                      ☐                      ☐                      ☐

Weiter

Abbildung A.9. Screenshot Onlinefragebogen der Version GV/GSV-Frauen



### 9. Fragen zur Körpergröße

Ich bin groß.

völlig unzutreffend	unzutreffend	weder noch	zutreffend	völlig zutreffend
1	2	3	4	5
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Ich halte mich für groß.

völlig unzutreffend	unzutreffend	weder noch	zutreffend	völlig zutreffend
1	2	3	4	5
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

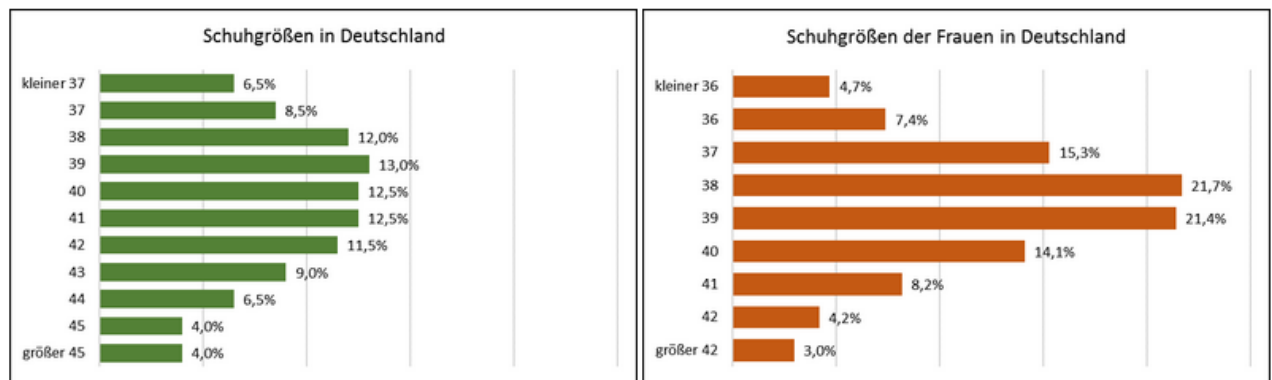
Andere halten mich für groß.

völlig unzutreffend	unzutreffend	weder noch	zutreffend	völlig zutreffend
1	2	3	4	5
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Weiter



Abbildung A.10. Screenshot Onlinefragebogen der Version GV/GSV-Frauen



#### 10. Fragen zur Schuhgröße

Ich habe große Füße.

völlig unzutreffend    unzutreffend    weder noch    zutreffend    völlig zutreffend

1                      2                      3                      4                      5

☐                      ☐                      ☐                      ☐                      ☐

Ich halte meine Füße für groß.

völlig unzutreffend    unzutreffend    weder noch    zutreffend    völlig zutreffend

1                      2                      3                      4                      5

☐                      ☐                      ☐                      ☐                      ☐

Andere halten meine Füße für groß.

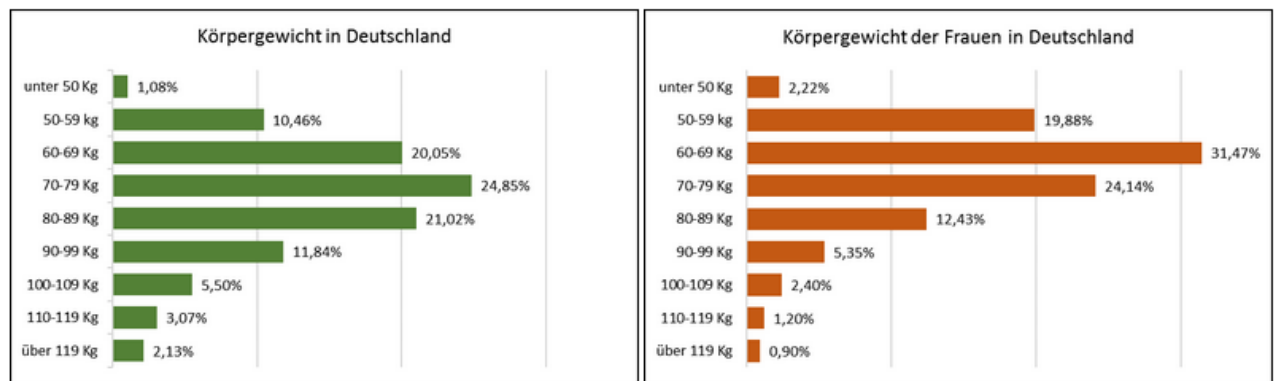
völlig unzutreffend    unzutreffend    weder noch    zutreffend    völlig zutreffend

1                      2                      3                      4                      5

☐                      ☐                      ☐                      ☐                      ☐

Weiter

Abbildung A.11. Screenshot Onlinefragebogen der Version GV/GSV-Frauen



#### 11. Fragen zum Gewicht

Ich wiege viel.

völlig unzutreffend	unzutreffend	weder noch	zutreffend	völlig zutreffend
1	2	3	4	5
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Ich finde, dass ich viel wiege.

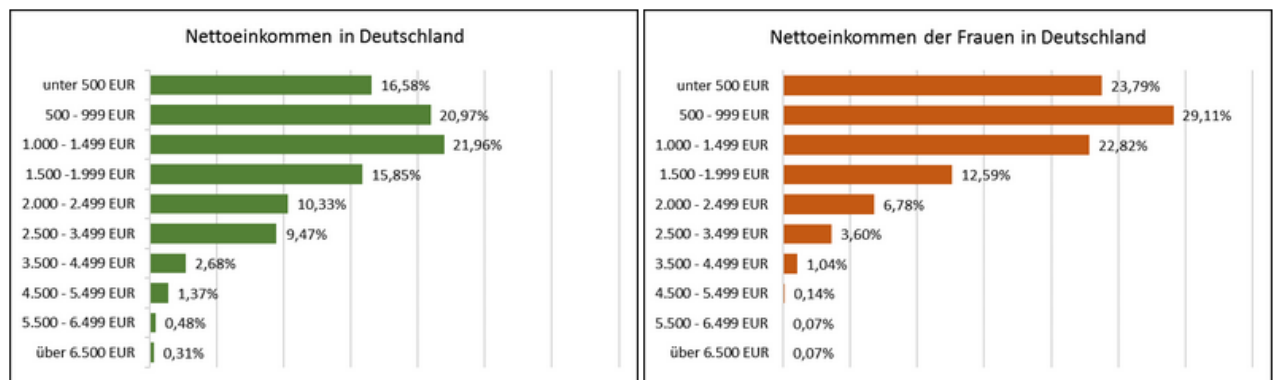
völlig unzutreffend	unzutreffend	weder noch	zutreffend	völlig zutreffend
1	2	3	4	5
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Andere finden, dass ich viel wiege.

völlig unzutreffend	unzutreffend	weder noch	zutreffend	völlig zutreffend
1	2	3	4	5
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Weiter

Abbildung A.12. Screenshot Onlinefragebogen der Version GV/GSV-Frauen



## 12. Fragen zum Einkommen

Ich verdiene viel.

völlig unzutreffend	unzutreffend	weder noch	zutreffend	völlig zutreffend
1	2	3	4	5
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Ich finde, dass ich viel verdiene.

völlig unzutreffend	unzutreffend	weder noch	zutreffend	völlig zutreffend
1	2	3	4	5
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Andere finden, dass ich viel verdiene.

völlig unzutreffend	unzutreffend	weder noch	zutreffend	völlig zutreffend
1	2	3	4	5
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Weiter

Abbildung A.13. Screenshot Onlinefragebogen der Version GV/GSV-Frauen

---

## **Vielen Dank für Ihre Teilnahme!**

Wir möchten uns ganz herzlich für Ihre Mithilfe bedanken.

Ihre Antworten wurden gespeichert, Sie können das Browser-Fenster nun schließen.

---

Kathryn Eichhorn, Ludwig-Maximilians-Universität München 2016

Tabelle A.2.

*Ergebnisse Likelihood Ratio Tests für Körpergewicht für Frauen*

Merkmal und Item-Nr.	GV <sup>1</sup>		GSV <sup>2</sup>		GV/GSV <sup>3</sup>		KV <sup>4</sup>	
	$\chi^2$	<i>p</i>	$\chi^2$	<i>p</i>	$\chi^2$	<i>p</i>	$\chi^2$	<i>p</i>
Körpergewicht (kg) <sup>5</sup>								
4	21.08	<b>&lt; .001</b>	25.06	<b>&lt; .001</b>	28.76	<b>&lt; .001</b>	16.19	<b>&lt; .001</b>
4(a)	64.27	<b>&lt; .001</b>	46.33	<b>&lt; .001</b>	46.89	<b>&lt; .001</b>	50.42	<b>&lt; .001</b>
4(b)	4.90	.08	4.54	.10	4.34	.11	3.26	.19
Körpergewicht (BMI)								
4	22.59	<b>&lt; .001</b>	30.19	<b>&lt; .001</b>	30.83	<b>&lt; .001</b>	17.79	<b>&lt; .001</b>
4(a)	64.66	<b>&lt; .001</b>	50.40	<b>&lt; .001</b>	48.28	<b>&lt; .001</b>	50.89	<b>&lt; .001</b>
4(b)	6.08	<b>.04</b>	6.50	<b>.03</b>	4.49	.10	5.64	<b>.05</b>

*Anmerkungen.*<sup>1</sup> Anzeige Balkendiagramm mit der Gesamtmerkmalsverteilung in der Population.<sup>2</sup> Anzeige Balkendiagramm mit der geschlechtsspezifischen Merkmalsverteilung in der Population.<sup>3</sup> Gleichzeitige Anzeige Balkendiagramme mit der Gesamtmerkmalsverteilung und der geschlechtsspezifischen Merkmalsverteilung in der Population.<sup>4</sup> Anzeige keiner Merkmalsverteilung. Bei der Bedingung KV handelt es sich um die Kontrollbedingung.<sup>5</sup> Die Ergebnisse für Körpergewicht in Kilogramm wurden aus Eichhorn (2016) entnommen.

(a) Egozentrische Perspektive; (b) Allozentrische Perspektive.

Signifikante Ergebnisse ( $p \leq .05$ ) sind fett gedruckt.

Tabelle A.3.

*Ergebnisse Likelihood Ratio Tests für Körpergewicht für Männer*

Merkmal und Item-Nr.	GV <sup>1</sup>		GSV <sup>2</sup>		GV/GSV <sup>3</sup>		KV <sup>4</sup>	
	$\chi^2$	<i>p</i>	$\chi^2$	<i>p</i>	$\chi^2$	<i>p</i>	$\chi^2$	<i>p</i>
Körpergewicht (kg) <sup>5</sup>								
4	7.58	<b>.02</b>	3.25	.19	10.90	<b>.004</b>	16.97	<b>&lt; .001</b>
4(a)	22.94	<b>&lt; .001</b>	15.46	<b>&lt; .001</b>	13.78	<b>.001</b>	14.93	<b>&lt; .001</b>
4(b)	1.94	.37	0.89	.63	3.49	.17	14.08	<b>&lt; .001</b>
Körpergewicht (BMI)								
4	8.70	<b>.01</b>	3.85	.14	10.79	<b>.004</b>	18.93	<b>&lt; .001</b>
4(a)	24.24	<b>&lt; .001</b>	14.88	<b>&lt; .001</b>	13.16	<b>.001</b>	16.16	<b>&lt; .001</b>
4(b)	2.90	.23	0.86	.64	3.39	.18	14.27	<b>&lt; .001</b>

*Anmerkungen.*<sup>1</sup> Anzeige Balkendiagramm mit der Gesamtmerkmalsverteilung in der Population.<sup>2</sup> Anzeige Balkendiagramm mit der geschlechtsspezifischen Merkmalsverteilung in der Population.<sup>3</sup> Gleichzeitige Anzeige Balkendiagramme mit der Gesamtmerkmalsverteilung und der geschlechtsspezifischen Merkmalsverteilung in der Population.<sup>4</sup> Anzeige keiner Merkmalsverteilung. Bei der Bedingung KV handelt es sich um die Kontrollbedingung.<sup>5</sup> Die Ergebnisse für Körpergewicht in Kilogramm wurden aus Eichhorn (2016) entnommen.

(a) Egozentrische Perspektive; (b) Allozentrische Perspektive.

Signifikante Ergebnisse ( $p \leq .05$ ) sind fett gedruckt.

## B. Anhang zu Studie II

Tabelle B.1.

*Items zur Selbstbewertung auf vier Merkmalsdimensionen*

Merkmal und Item-Nummer		Item
Alter	1(a)	Wie schätzen <b>Sie</b> Ihr Alter ein?
	1(b)	Wie schätzen <b>andere</b> Ihr Alter ein?
Einkommen	2(a)	Wie schätzen <b>Sie</b> Ihr Einkommen ein?
	2(b)	Wie schätzen <b>andere</b> Ihr Einkommen ein?
Körpergröße	3(a)	Wie schätzen <b>Sie</b> Ihre Körpergröße ein?
	3(b)	Wie schätzen <b>andere</b> Ihre Körpergröße ein?
Körpergewicht	4(a)	Wie schätzen <b>Sie</b> Ihr Körpergewicht ein?
	4(b)	Wie schätzen <b>andere</b> Ihr Körpergewicht ein?

*Anmerkungen.*

Die Reihenfolge der Merkmale entspricht der Abfragereihenfolge im Fragebogen.

(a) Egozentrische Perspektive; (b) Allozentrische Perspektive.

Abbildung B.1. Screenshot Onlinefragebogen der Version 4PS



### Informationen zur Befragung

Sehr geehrte Teilnehmerin, sehr geehrter Teilnehmer,

bitte lesen Sie die folgenden Informationen zur Befragung und zum Schutz der erhobenen Daten durch.

Die Studie dauert ca. **2 Minuten**.

Ziel der Befragung ist die grundlagenwissenschaftliche Untersuchung psychologischer Messinstrumente im Rahmen des Promotionsstudiums der Versuchsleiterin im Fach Psychologie an der Ludwig-Maximilians-Universität München.

#### 1. Daten

Alle erhobenen Daten werden anonym behandelt. Ihr Name wird an keiner Stelle erfasst. Rückschlüsse auf Ihre Person sind nicht möglich und nicht von Interesse. Die aufgezeichneten Daten lassen sich in folgende Kategorien unterteilen:

- a) Demografische Daten
- b) Antworten auf Fragebogenitems

Sie werden zunächst um die Angabe **konkreter Werte** gebeten. Anschließend sollen Sie Einschätzungen auf einer vorgegebenen Skala abgeben. Sie werden auch nach Ihrem Einkommen gefragt. Diese konkreten Einkommenswerte werden **nicht** separat analysiert. Wesentlich sind nur Ihre später gegebenen Einschätzungen.

#### 2. Aufbewahrung und Löschung der Daten

Die anonymisierten Daten (Demografische Daten, Antworten auf Fragebogenitems) werden auf Datenträgern der Versuchsleiterin gespeichert und sind dabei nicht öffentlich zugänglich.

Die Verwendung erfolgt im Rahmen des Promotionsstudiums und zur Erstellung der Dissertation der Versuchsleiterin. Die VersuchsteilnehmerInnen können jederzeit und ohne Nennung von Gründen die Befragung abbrechen.

**Wenn Sie mit den oben genannten Punkten zur Datenerhebung einverstanden sind, geht es nun weiter zum Fragebogen. Sollten Sie mit den oben genannten Punkten nicht einverstanden sein, können Sie die Befragung an dieser Stelle beenden.**

Gerne können Sie Fragen zur Studie an die Versuchsleiterin richten:

Kathryn.Eichhorn@psy.lmu.de

Weiter

Abbildung B.2. Screenshot Onlinefragebogen der Version 4PS



**1. Wie alt sind Sie?**

Ich bin  Jahre alt.

**2. Welche Schulbildung haben Sie?**

- ☐ keine allgemeine Hochschulreife  
☐ allgemeine Hochschulreife (allgemeines Abitur)

**3. Wie groß sind Sie (in cm)?**

Ich bin  cm groß.

**4. Wieviel wiegen Sie (in kg)?**

Ich wiege  kg.

**5. Wie hoch ist Ihr monatliches Nettoeinkommen (in Euro)?**

Mein monatliches Nettoeinkommen beträgt  EUR.


**6. Welches Geschlecht haben Sie?**

- ☐ weiblich  
☐ männlich

Weiter



Abbildung B.3. Screenshot Onlinefragebogen der Version 4PS



LUDWIG-  
MAXIMILIANS-  
UNIVERSITÄT  
MÜNCHEN

**Fragen zum Alter**  
Wählen Sie für Ihre Einschätzung einen Wert auf der Skala von „niedrig“ bis „hoch“.  
Zwischenwerte können entsprechend gewählt werden.

Wie schätzen **Sie** Ihr Alter ein?

niedrig

☐

☐

☐

☐

hoch

Wie schätzen **andere** Ihr Alter ein?

niedrig

☐

☐

☐

☐

hoch


Weiter

---

Kathryn Eichhorn, Ludwig-Maximilians-Universität München 2017

33% ausgefüllt

Abbildung B.4. Screenshot Onlinefragebogen der Version 4PS



LUDWIG-  
MAXIMILIANS-  
UNIVERSITÄT  
MÜNCHEN

**Fragen zum monatlichen Nettoeinkommen**  
Wählen Sie für Ihre Einschätzung einen Wert auf der Skala von „niedrig“ bis „hoch“.  
Zwischenwerte können entsprechend gewählt werden.

Wie schätzen **Sie** Ihr Einkommen ein?

niedrig

☐

☐

☐

☐

hoch

Wie schätzen **andere** Ihr Einkommen ein?

niedrig

☐

☐

☐

☐

hoch


Weiter

---

Kathryn Eichhorn, Ludwig-Maximilians-Universität München 2017

50% ausgefüllt

Abbildung B.5. Screenshot Onlinefragebogen der Version 4PS



LUDWIG-  
MAXIMILIANS-  
UNIVERSITÄT  
MÜNCHEN

**Fragen zur Körpergröße**  
Wählen Sie für Ihre Einschätzung einen Wert auf der Skala von „niedrig“ bis „hoch“. Zwischenwerte können entsprechend gewählt werden.

Wie schätzen **Sie** Ihre Körpergröße ein?

niedrig

☐

☐

☐

☐

hoch

Wie schätzen **andere** Ihre Körpergröße ein?

niedrig

☐

☐

☐

☐

hoch


Weiter

---

Kathryn Eichhorn, Ludwig-Maximilians-Universität München 2017

67% ausgefüllt

Abbildung B.6. Screenshot Onlinefragebogen der Version 4PS



LUDWIG-  
MAXIMILIANS-  
UNIVERSITÄT  
MÜNCHEN

**Fragen zum Körpergewicht**  
Wählen Sie für Ihre Einschätzung einen Wert auf der Skala von „niedrig“ bis „hoch“. Zwischenwerte können entsprechend gewählt werden.

Wie schätzen **Sie** Ihr Körpergewicht ein?

niedrig

☐

☐

☐

☐

hoch

Wie schätzen **andere** Ihr Körpergewicht ein?

niedrig

☐

☐

☐

☐

hoch

Weiter

---

Kathryn Eichhorn, Ludwig-Maximilians-Universität München 2017

83% ausgefüllt

Abbildung B.7. Screenshot Onlinefragebogen der Version 4PS

## **Vielen Dank für Ihre Teilnahme!**

Wir möchten uns ganz herzlich für Ihre Mithilfe bedanken.

Ihre Antworten wurden gespeichert, Sie können das Browser-Fenster nun schließen.

---

Kathryn Eichhorn, Ludwig-Maximilians-Universität München 2017

Abbildung B.8. Histogramme der realen Werte für die Merkmale Alter und Einkommen

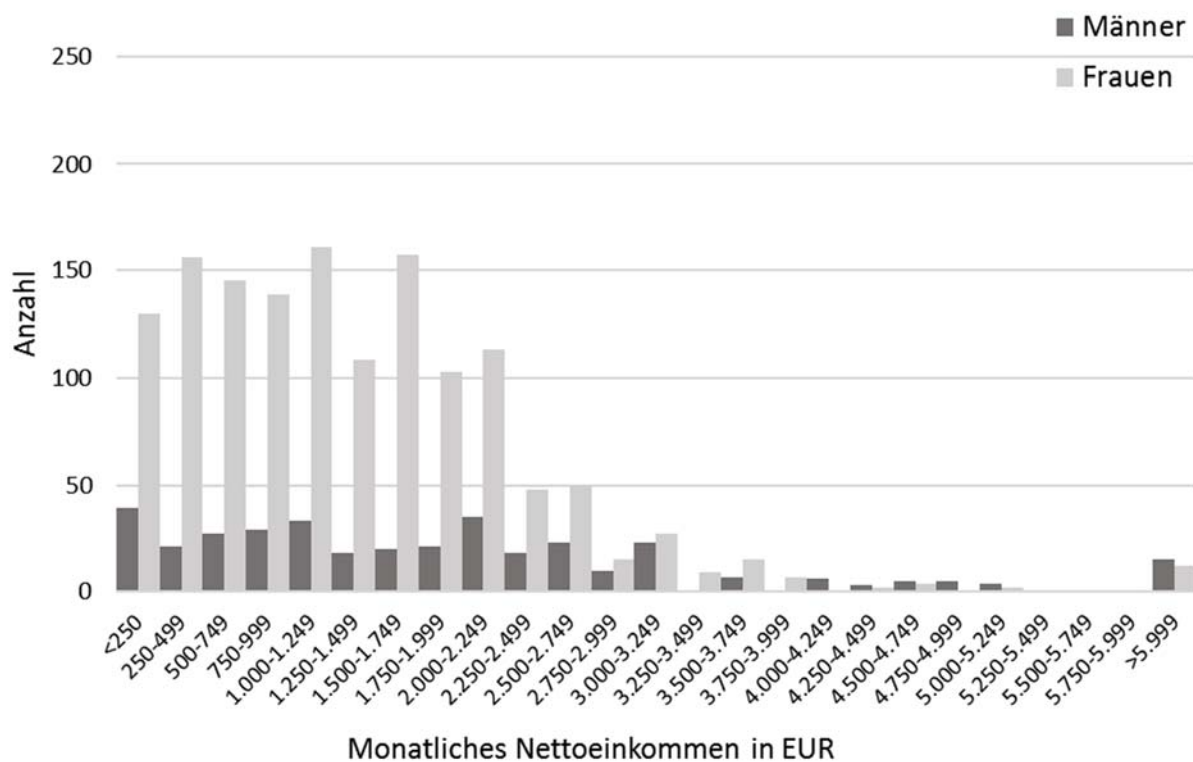
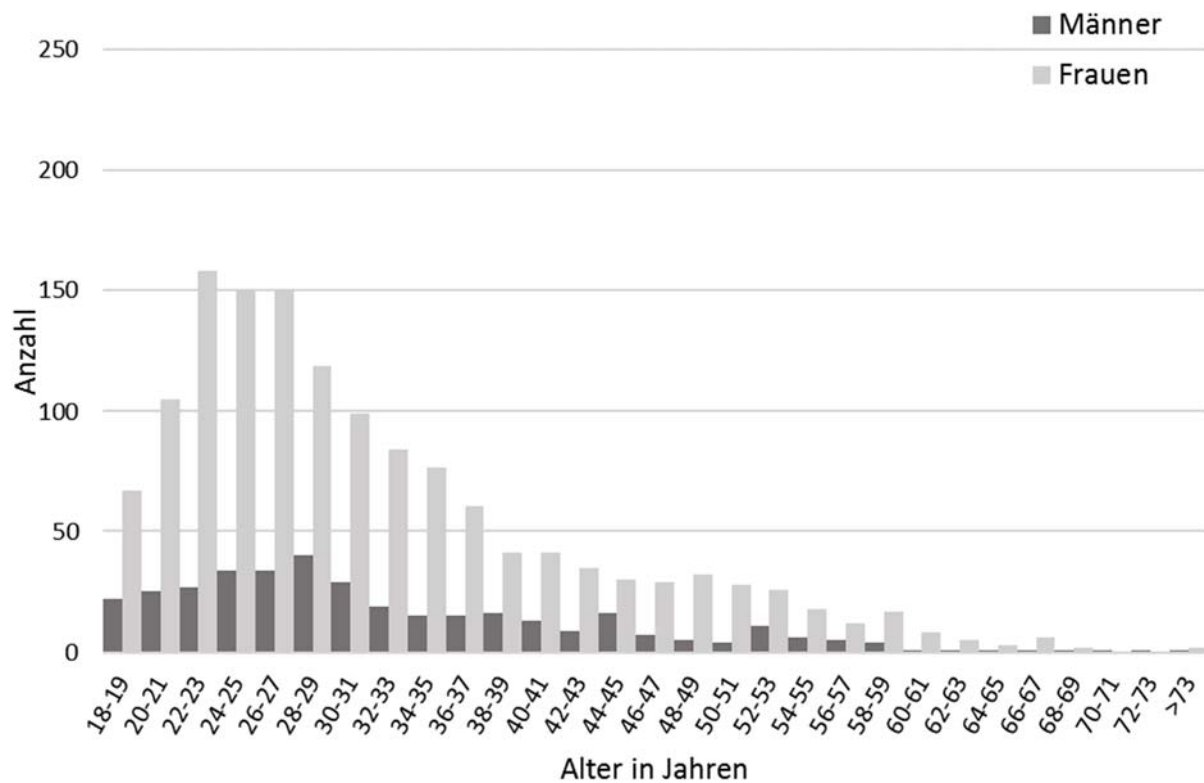


Abbildung B.9. Histogramme der realen Werte für die Merkmale Körpergewicht und -größe

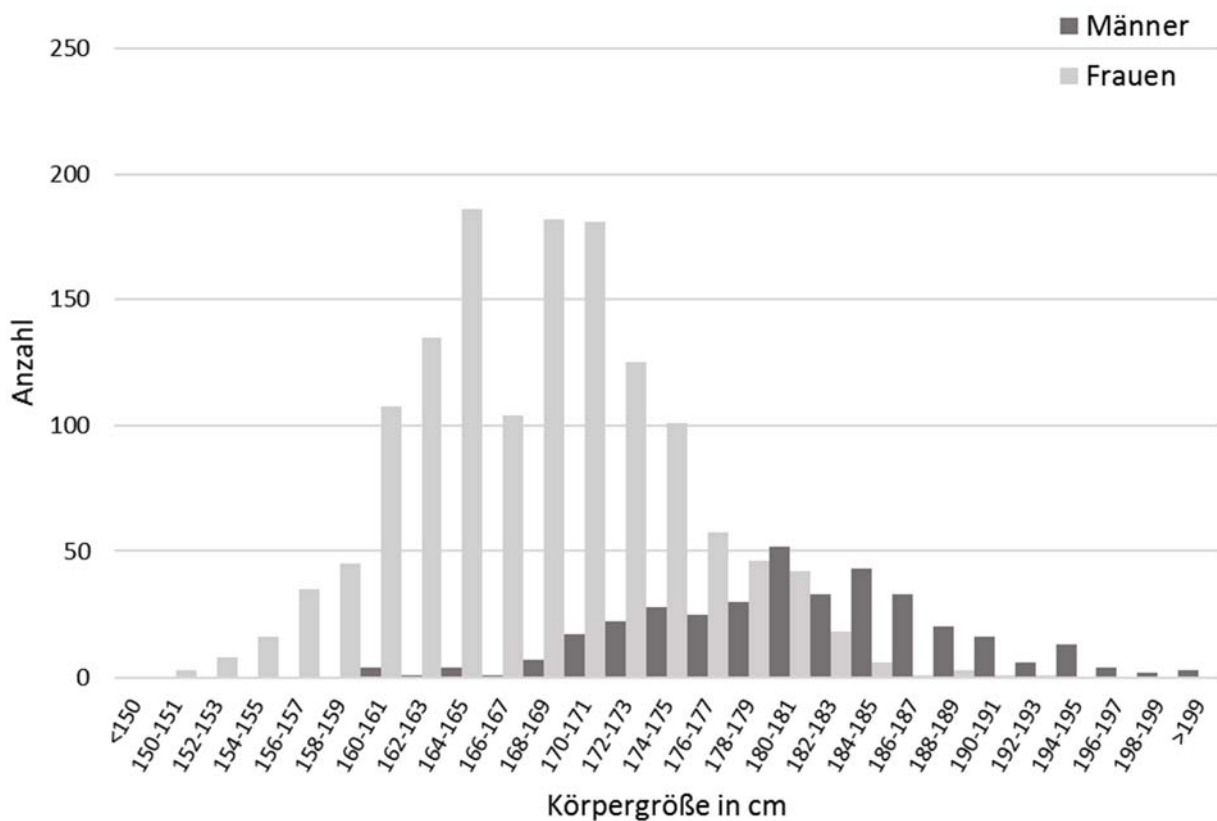
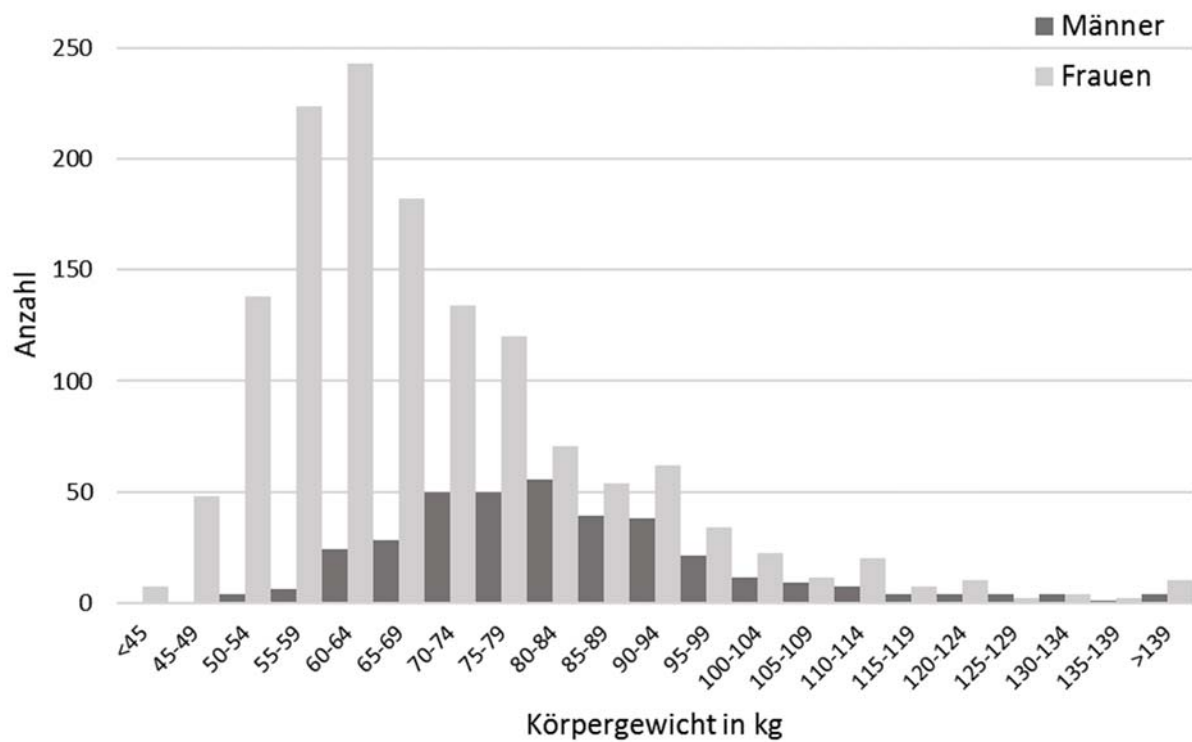


Tabelle B.2.

*Ergebnisse Likelihood Ratio Tests der Bedingungen 5PS<sup>1</sup> und 4PS<sup>2</sup> für Frauen und Männer*

Merkmal und Item-Nr.		Frauen				Männer			
		5PS <sup>1</sup>		4PS <sup>2</sup>		5PS <sup>1</sup>		4PS <sup>2</sup>	
		$\chi^2$	<i>p</i>	$\chi^2$	<i>p</i>	$\chi^2$	<i>p</i>	$\chi^2$	<i>p</i>
Alter									
	1(a)	8.67	<b>.01</b>	8.40	<b>.003</b>	0.01	.99	0	1
	1(b)	6.42	<b>.04</b>	7.51	<b>.006</b>	3.32	.18	1.57	.21
Einkommen									
	2(a)	10.88	<b>.004</b>	2.62	.11	2.92	.23	0.78	.38
	2(b)	9.22	<b>.01</b>	14.25	<b>&lt; .001</b>	7.37	<b>.03</b>	0.05	.83
Körpergröße									
	3(a)	13.02	<b>.001</b>	0.02	.88	0.98	.61	1.47	.23
	3(b)	9.96	<b>.006</b>	2.79	.09	0.10	.95	0.57	.45
Körpergewicht									
	4(a)	2.11	.35	0.33	.57	1.71	.42	1.47	.22
	4(b)	3.04	.22	4.08	<b>.04</b>	1.29	.52	0.95	.33

*Anmerkungen.*<sup>1</sup> Fünfstufige Antwortskala.<sup>2</sup> Vierstufige Antwortskala.

(a) Egozentrische Perspektive; (b) Allozentrische Perspektive.

Signifikante Ergebnisse ( $p \leq .05$ ) sind fett gedruckt.



Tabelle B.3.

*Ergebnisse Brant-Tests der Bedingungen 5PS<sup>1</sup> und 4PS<sup>2</sup> für Frauen und Männer*

Merkmal und Item-Nr.		Frauen				Männer			
		5PS <sup>1</sup>		4PS <sup>2</sup>		5PS <sup>1</sup>		4PS <sup>2</sup>	
		$\chi^2$	<i>p</i>	$\chi^2$	<i>p</i>	$\chi^2$	<i>p</i>	$\chi^2$	<i>p</i>
Alter									
	1(a)	15.29	<b>&lt; .001</b>	1.21	.55	3.12	.37	-	-
	1(b)	13.19	<b>&lt; .001</b>	3.00	.22	5.09	.17	0.66	.72
Einkommen									
	2(a)	24.05	<b>&lt; .001</b>	25.16	<b>&lt; .001</b>	47.94	<b>&lt; .001</b>	10.84	<b>&lt; .001</b>
	2(b)	11.19	<b>.01</b>	21.25	<b>&lt; .001</b>	1.90	.59	6.38	<b>.04</b>
Körpergröße									
	3(a)	24.79	<b>&lt; .001</b>	0.35	.84	5.50	.14	3.26	.20
	3(b)	0.61	.89	0.37	.83	1.70	.64	0.42	.81
Körpergewicht									
	4(a)	20.54	<b>&lt; .001</b>	12.8	<b>&lt; .001</b>	5.77	.12	0.47	.79
	4(b)	6.53	.09	17.06	<b>&lt; .001</b>	4.81	.19	0.94	.62

*Anmerkungen.*<sup>1</sup> Fünfstufige Antwortskala.<sup>2</sup> Vierstufige Antwortskala.

(a) Egozentrische Perspektive; (b) Allozentrische Perspektive.

Signifikante Ergebnisse ( $p \leq .05$ ) sind fett gedruckt.

## C. Anhang zu Studie III

### Abbildung C.1. Fragebogen Prototypenansatz

#### Informationen zur Befragung



Sehr geehrte Teilnehmerin, sehr geehrter Teilnehmer,

bitte lesen Sie die folgenden Informationen zur Befragung und zum Schutz der erhobenen Daten durch.

Die Befragung dauert ca. **10 Minuten**.

#### 1. Ziel der Befragung

Ziel der Befragung ist die grundlagenwissenschaftliche Untersuchung psychologischer Messinstrumente im Rahmen des Promotionsstudiums der Versuchsleiterin am Lehrstuhl für psychologische Methodenlehre und Diagnostik an der Ludwig-Maximilians-Universität München. Die Teilnahme ist freiwillig und ohne Vergütung. Die Befragung kann jederzeit ohne Nennung von Gründen und ohne Konsequenzen abgebrochen werden.

#### 2. Daten

Bei der Befragung werden Sie gebeten Ihre prototypischen Vorstellungen für die Merkmale Alter, Körpergröße und Körpergewicht in Stichpunkten darzulegen. Die aufgezeichneten Daten lassen sich in folgende Kategorien unterteilen:

- a) Demographische Daten
- b) Befragung

Alle erhobenen Daten werden anonym behandelt. Ihr Name wird an keiner Stelle erfasst. Rückschlüsse auf Ihre Person sind nicht möglich und nicht von Interesse.

#### 3. Aufbewahrung und Löschung der Daten

Die demographischen Daten und die prototypischen Vorstellungen werden abgeschrieben und auf Datenträgern der Versuchsleiterin gespeichert.

Die Daten und Ergebnisse dieser Befragung werden im Rahmen des Promotionsstudiums und zur Erstellung der Dissertation der Versuchsleiterin verwendet und anschließend ggf. als wissenschaftliche Publikation veröffentlicht. Dies geschieht in anonymisierter Form, d. h., ohne dass die Daten einer spezifischen Person zugeordnet werden können. Die Ursprungsfragebögen werden nicht veröffentlicht und bleiben in Händen der Versuchsleiterin. Dadurch wird sichergestellt, dass Rückschlüsse auf einzelne Personen, beispielsweise anhand einer unverkennbaren Handschrift nicht möglich sind.

**Wenn Sie mit den oben genannten Punkten zur Datenerhebung einverstanden sind, geht es nun weiter zur Befragung. Sollten Sie mit den oben genannten Punkten nicht einverstanden sein, können Sie die Befragung an dieser Stelle beenden.**

Bei Fragen zur Studie können Sie gerne eine E-Mail schreiben an:

Kathryn.Eichhorn@psy.lmu.de

## Abbildung C.2. Fragebogen Prototypenansatz

### a) Demographische Daten

Geschlecht: ☐ weiblich ☐ männlich  
Bildungsniveau: ☐ Abitur ☐ kein Abitur  
Alter in Jahren: \_\_\_\_\_

### b) Befragung

#### 1. Alter

Denken Sie bitte an eine Ihnen gut bekannte Person, die ein **hohes Alter** hat. Welche **Verhaltensweisen** zeigt diese Person, die **besonders typisch für ihr hohes Alter** sind. Beschreiben Sie diese Verhaltensweisen in kurzen Stichpunkten:

---

---

---

---

---

---

---

---

---

---

---

Denken Sie nun bitte an eine Ihnen gut bekannte erwachsene Person, die **mindestens 18 Jahre alt** ist (junger Erwachsene/r). Welche **Verhaltensweisen** zeigt diese Person, die **besonders typisch für ihr junges Alter** sind. Beschreiben Sie diese Verhaltensweisen in kurzen Stichpunkten:

---

---

---

---

---

---

---

---

---

---

---

## Abbildung C.3. Fragebogen Prototypenansatz

### 2. Körpergröße

Denken Sie bitte an eine Ihnen gut bekannte Person, die **sehr groß** ist. Welche **Verhaltensweisen** zeigt diese Person, die **besonders typisch für ihre große Körpergröße** sind. Beschreiben Sie diese Verhaltensweisen in kurzen Stichpunkten:

---

---

---

---

---

---

---

---

---

---

---

Denken Sie nun bitte an eine Ihnen gut bekannte erwachsene Person, die **besonders klein** ist. Welche **Verhaltensweisen** zeigt diese Person, die **besonders typisch für ihre kleine Körpergröße** sind. Beschreiben Sie diese Verhaltensweisen in kurzen Stichpunkten:

---

---

---

---

---

---

---

---

---

---

---

## Abbildung C.4. Fragebogen Prototypenansatz

### 3. Körpergewicht

Denken Sie bitte an eine Ihnen gut bekannte Person, die ein **hohes Körpergewicht** hat. Welche **Verhaltensweisen** zeigt diese Person, die **besonders typisch für ihr hohes Körpergewicht** sind. Beschreiben Sie diese Verhaltensweisen in kurzen Stichpunkten:

---

---

---

---

---

---

---

---

---

---

Denken Sie nun bitte an eine Ihnen gut bekannte Person, die ein besonders **niedriges Körpergewicht** hat. Welche **Verhaltensweisen** zeigt diese Person, die **besonders typisch für ihr niedriges Körpergewicht** sind. Beschreiben Sie diese Verhaltensweisen in kurzen Stichpunkten:

---

---

---

---

---

---

---

---

---

---

Tabelle C.1.

*Items für die Merkmale Alter, Körpergröße und Körpergewicht – Pilottest*

Merkmal und Item-Nummer		Item
Alter		
1p	(D, P)	Ich habe körperliche Gebrechen.
2p	(D)	Im Laufe der Zeit hat meine geistige Leistungsfähigkeit abgenommen.
3p	(D, P)	Ich habe viel Lebenserfahrung.
4p	(D, P)	Aufgrund meiner Lebenserfahrung werde ich oft um Rat gefragt.
5p	(D, P)	Im Laufe der Zeit hat meine Reaktionsfähigkeit abgenommen.
6p	(D, P)	Im Laufe der Zeit hat meine Merkfähigkeit abgenommen.
7p	(D, P)	Ich ertappe mich oft dabei, dass ich dieselben Geschichten mehrmals erzähle.
8p	(D, P)	Ich mache mir viele Gedanken über das Älterwerden.
9p	(D)	Den größten Teil meines Lebens habe ich bereits gelebt.
10p	(D, P)	Ich habe zunehmend Schwierigkeiten technischen Weiterentwicklungen zu folgen.
11p	(D, P)	Bei der Bedienung von Computern oder Smartphones frage ich oft nach Hilfe.
12p	(D)	Ich habe jahrelange Berufserfahrung.
13p	(D)*	Nach einer langen Partynacht bin ich schnell wieder erholt.
14p	(D, P)*	Mein ganzes Leben liegt noch vor mir.
15p	(D, P)	Früher war ich risikofreudiger.
16p	(P)	Im Vergleich zu früher bin ich vergesslicher geworden.
17p	(P)	Ich verrichte Alltagstätigkeiten langsamer als früher.
18p	(P)	Ich schwelge oft in alten Erinnerungen.
19p	(P)	Im Vergleich zu früher lege ich heute mehr Wert auf einen geregelten Tagesablauf.
20p	(P)	Ich nehme gerne an organisierten Tagesausflügen teil.
21p	(P)*	Ich muss mich selbst erst noch finden.
Körpergröße		
22p	(D, P)	Ich ziehe den Kopf ein, wenn ich durch Türen gehe.
23p	(D, P)	Wenn ich mich mit anderen Menschen unterhalte, muss ich nach unten schauen.
24p	(D)	Inmitten einer Menschenmenge kann ich über die meisten anderen Menschen hinwegblicken.
25p	(D)	Stühle und Tische sind meistens zu niedrig für mich.
26p	(D, P)	Im Supermarkt komme ich ohne Probleme an die Sachen im obersten Regal.
27p	(D, P)	Meinen Autositz habe ich ganz nach hinten eingestellt.
28p	(D, P)	Wenn ich ein Auto fahre, das vorher jemand anderes gefahren hat, muss ich üblicherweise den Sitz nach hinten schieben.
29p	(D, P)	Ich mache mich in meiner Körperhaltung oft kleiner.
30p	(D, P)	Ich würde mich gut in einem Basketballteam machen.
31p	(D)	Wenn ich mich in Hotelbetten gerade ausstrecke, dann stehen meine Füße über.
32p	(D, P)	Wenn ich im Flugzeug sitze, stoße ich mit den Knien am Vordersitz an.

Tabelle C.1. (Fortsetzung)

*Items für die Merkmale Alter, Körpergröße und Körpergewicht – Pilottest*

Merkmal und Item-Nummer		Item
33p	(D, P)	Hosen sind mir häufig zu kurz.
34p	(D, P)	Bei Gruppenfotos stehe ich hinten.
35p	(D, P)*	Wenn ich im Flugzeug sitze, kann ich meine Beine ausstrecken.
36p	(D, P)*	Wenn ich mir Hosen kaufe, muss ich üblicherweise den Saum kürzen.
37p	(D, P)*	Ich brauche einen Stuhl, wenn ich Sachen aus dem obersten Regal holen möchte.
38p	(D, P)*	Im Supermarkt habe ich Probleme, an die Sachen im obersten Regal zu kommen.
39p	(D, P)*	Wenn ich mich mit anderen Menschen unterhalte, muss ich nach oben schauen.
40p	(D, P)*	Wenn ich ein Auto fahre, das vorher jemand anderes gefahren hat, muss ich üblicherweise den Sitz nach vorne schieben.
41p	(D, P)*	Ich muss mich bei Gruppenfotos nach vorne stellen, damit ich gut zu sehen bin.
42p	(P)	Wenn ich andere Menschen zur Begrüßung umarme, muss ich mich nach unten beugen.
43p	(P)*	Wenn ich andere Menschen zur Begrüßung umarme, muss ich mich nach oben strecken.
<b>Körpergewicht</b>		
44p	(D, P)	Beim Treppensteigen komme ich schnell außer Atem und ins Schwitzen.
45p	(D)	Zwischen den Armlehnen in Flugzeugsitzen habe ich keinen Platz.
46p	(D)	Ich habe oft Angst, dass Stühle unter mir nachgeben.
47p	(D)	Möglicherweise denken andere Menschen von mir, dass ich abends auf dem Sofa sitze und Schokolade und Chips esse.
48p	(D, P)	Bei Hosen habe ich eine große Bundweite.
49p	(D)	Ich bin dick.
50p	(D)	Ich wiege viel.
51p	(D, P)	Ich müsste abnehmen.
52p	(D, P)*	Bei Hosen habe ich eine kleine Bundweite.
53p	(D)*	Ich kann mich gut durch enge Spalten quetschen.
54p	(D)*	Zwischen den Armlehnen in Flugzeugsitzen habe ich viel Platz.
55p	(D, P)*	Wenn ich die Wahl habe zwischen Aufzug und Treppe, dann nehme ich die Treppe.
56p	(D)*	Ich bin schlank.
57p	(D)*	Ich wiege wenig.
58p	(D)*	Im Aufzug nehme ich wenig Platz ein.
59p	(D)	Im Aufzug nehme ich viel Platz ein.
60p	(P)	Wenn ich die Wahl habe zwischen Aufzug und Treppe, dann nehme ich den Aufzug.
61p	(P)	Ich sollte weniger essen.

*Anmerkungen.*

p = Item-Nummer für den Pilottest; D = Items hervorgegangen aus dem deduktiven Testkonstruktionsansatz; P = Items hervorgegangen aus dem Prototypenansatz.

\* Items mit negativer Polung.

Abbildung C.5. Exemplarischer Auszug Onlinefragebogen Pilottest



Wählen Sie für Ihre Einschätzung einen Wert auf der Skala von „trifft nicht zu“ bis „trifft zu“. Zwischenwerte können entsprechend gewählt werden.

**Ich ziehe den Kopf ein, wenn ich durch Türen gehe.**

trifft nicht zu						trifft zu
1	2	3	4	5	6	
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	

**Beim Treppensteigen komme ich schnell außer Atem und ins Schwitzen.**

trifft nicht zu						trifft zu
1	2	3	4	5	6	
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	

**Ich habe viel Lebenserfahrung.**

trifft nicht zu						trifft zu
1	2	3	4	5	6	
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	

**Wenn ich mich mit anderen Menschen unterhalte, muss ich nach unten schauen.**

trifft nicht zu						trifft zu
1	2	3	4	5	6	
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	

**Bei Hosen habe ich eine große Bundweite.**

trifft nicht zu						trifft zu
1	2	3	4	5	6	
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	



Tabelle C.2.

*Mittelwerte (M), Standardabweichungen (SD) und Trennschärfen ( $r_{i(t-i)}$ ) – Pilottest*

Merkmal und Item-Nummer		<i>M</i>	<i>SD</i>	$r_{i(t-i)}$
Alter				
1p	(D, P)	2.47	1.51	.34
2p	(D)	2.50	1.38	.66
3p	(D, P)	3.71	1.26	.39
4p	(D, P)	3.38	1.40	.31
5p	(D, P)	2.31	1.26	.62
6p	(D, P)	2.95	1.48	.63
7p	(D, P)	3.14	1.47	.11
8p	(D, P)	3.13	1.41	.32
9p	(D)	2.32	1.35	.59
10p	(D, P)	2.12	1.30	.43
11p	(D, P)	1.86	1.12	.28
12p	(D)	3.10	1.92	.48
13p	(D)*	3.88	1.54	.30
14p	(D, P)*	3.28	1.44	.49
15p	(D, P)	3.43	1.63	.48
16p	(P)	2.99	1.60	.53
17p	(P)	2.22	1.40	.53
18p	(P)	3.61	1.46	.29
19p	(P)	3.36	1.61	.28
20p	(P)	2.60	1.55	-.11
21p	(P)*	3.85	1.61	-.02
Körpergröße				
22p	(D, P)	1.62	1.19	.50
23p	(D, P)	1.99	1.28	.50
24p	(D)	2.20	1.50	.76
25p	(D)	1.70	1.18	.62
26p	(D, P)	3.69	1.87	.78
27p	(D, P)	1.80	1.36	.53
28p	(D, P)	2.36	1.64	.74
29p	(D, P)	2.37	1.47	.32
30p	(D, P)	1.65	1.20	.40
31p	(D)	1.69	1.31	.66
32p	(D, P)	2.53	1.71	.72
33p	(D, P)	2.10	1.65	.63
34p	(D, P)	3.36	1.72	.65

Tabelle C.2. (Fortsetzung)

*Mittelwerte (M), Standardabweichungen (SD) und Trennschärfen ( $r_{i(t-i)}$ ) – Pilottest*

Merkmal und Item-Nummer	<i>M</i>	<i>SD</i>	$r_{i(t-i)}$
35p (D, P)*	3.87	1.72	.62
36p (D, P)*	4.70	1.76	.55
37p (D, P)*	3.77	1.92	.69
38p (D, P)*	4.31	1.79	.76
39p (D, P)*	4.21	1.57	.68
40p (D, P)*	3.57	1.91	.73
41p (D, P)*	4.19	1.81	.76
42p (P)	2.21	1.50	.81
43p (P)*	4.17	1.63	.72
Körpergewicht			
44p (D, P)	2.72	1.41	.30
45p (D)	1.98	1.40	.43
46p (D)	1.32	0.84	.45
47p (D)	2.68	1.68	.40
48p (D, P)	2.25	1.47	.75
49p (D)	2.37	1.65	.82
50p (D)	2.69	1.72	.84
51p (D, P)	3.29	1.86	.76
52p (D, P)*	3.76	1.72	.68
53p (D)*	3.11	1.54	.65
54p (D)*	3.50	1.49	.44
55p (D, P)*	2.99	1.59	.29
56p (D)*	3.32	1.72	.79
57p (D)*	4.42	1.60	.56
58p (D)*	2.64	1.52	.64
59p (D)	1.89	1.24	.71
60p (P)	2.75	1.67	.34
61p (P)	3.24	1.71	.61

*Anmerkungen.**M* = Mittelwert; *SD* = Standardabweichung;  $r_{i(t-i)}$  = Trennschärfe.

p = Item-Nummer für den Pilottest.

Die Nummerierung der Items entspricht der Nummerierung in Tabelle C.1.

D = Items hervorgegangen aus dem deduktiven Testkonstruktionsansatz.

P = Items hervorgegangen aus dem Prototypenansatz.

\* Items mit negativer Polung.

Tabelle C.3.

*Explorative Faktorenanalyse (Maximum Likelihood, Promax) – Pilottest*

Merkmal und Item-Nummer		Faktor 1	Faktor 2	Faktor 3	$h^2$
Alter					
1p	(D, P)		.13	<b>.31</b>	.1424
2p	(D)		-.11	<b>.87</b>	.7051
3p	(D, P)		.13	<b>.22</b>	.0880
4p	(D, P)			<b>.15</b>	.0318
5p	(D, P)			<b>.74</b>	.5753
6p	(D, P)		-.16	<b>.90</b>	.7381
7p	(D, P)			<b>.20</b>	.0561
8p	(D, P)			<b>.35</b>	.1264
9p	(D)			<b>.52</b>	.2657
10p	(D, P)			<b>.42</b>	.2072
11p	(D, P)			<b>.24</b>	.0600
12p	(D)			<b>.32</b>	.1258
13p	(D)*	-.13		<b>.33</b>	.1131
14p	(D, P)*	-.10	.14	<b>.37</b>	.1947
15p	(D, P)			<b>.41</b>	.1882
16p	(P)		-.14	<b>.83</b>	.6235
17p	(P)			<b>.59</b>	.3936
18p	(P)		.10	<b>.24</b>	.0836
19p	(P)			<b>.26</b>	.0628
20p	(P)				.0059
21p	(P)*			<b>-.14</b>	.0168
Körpergröße					
22p	(D, P)	<b>.50</b>			.2481
23p	(D, P)	<b>.50</b>		.10	.2494
24p	(D)	<b>.79</b>			.6111
25p	(D)	<b>.62</b>		.10	.3928
26p	(D, P)	<b>.84</b>	-.16		.6718
27p	(D, P)	<b>.54</b>			.3018
28p	(D, P)	<b>.77</b>			.5928
29p	(D, P)	<b>.32</b>		.23	.1385
30p	(D, P)	<b>.42</b>		-.12	.1905
31p	(D)	<b>.67</b>			.4489
32p	(D, P)	<b>.72</b>	.11		.5633
33p	(D, P)	<b>.64</b>			.4063
34p	(D, P)	<b>.64</b>	.12		.4639

Tabelle C.3. (Fortsetzung)

*Explorative Faktorenanalyse (Maximum Likelihood, Promax) – Pilottest*

Merkmal und Item-Nummer	Faktor 1	Faktor 2	Faktor 3	$h^2$
35p (D, P)*	<b>.63</b>	.12		.4327
36p (D, P)*	<b>.59</b>			.3423
37p (D, P)*	<b>.74</b>	-.12	-.12	.5503
38p (D, P)*	<b>.80</b>			.6345
39p (D, P)*	<b>.71</b>			.4970
40p (D, P)*	<b>.76</b>			.5794
41p (D, P)*	<b>.78</b>			.6136
42p (P)	<b>.84</b>			.6876
43p (P)*	<b>.76</b>			.5682
Körpergewicht				
44p (D, P)		.17	<b>.26</b>	.1237
45p (D)	.23	<b>.39</b>		.2518
46p (D)		<b>.43</b>		.2195
47p (D)		<b>.39</b>		.1668
48p (D, P)		<b>.78</b>		.5958
49p (D)	-.14	<b>.93</b>		.8143
50p (D)		<b>.94</b>		.8465
51p (D, P)	-.11	<b>.86</b>		.7016
52p (D, P)*		<b>.72</b>		.4977
53p (D)*	.10	<b>.68</b>	-.10	.4623
54p (D)*	.24	<b>.45</b>		.2839
55p (D, P)*		<b>.17</b>	.16	.0739
56p (D)*		<b>.89</b>	-.12	.7049
57p (D)*		<b>.70</b>	-.21	.4159
58p (D)*	.12	<b>.66</b>		.4514
59p (D)		<b>.72</b>		.5543
60p (P)		<b>.21</b>	.18	.1047
61p (P)	-.12	<b>.66</b>		.4632

**Anmerkungen.** $h^2$  = Kommunalität des Items. Ladungen mit Betrag < .10 wurden unterdrückt.

Die stärkste Ladung für jedes Item ist fett gedruckt.

p = Item-Nummer für den Pilottest.

Die Nummerierung der Items entspricht der Nummerierung in Tabelle C.1.

D = Items hervorgegangen aus dem deduktiven Testkonstruktionsansatz.

P = Items hervorgegangen aus dem Prototypenansatz.

\* Items mit negativer Polung.

Tabelle C.4.

*Faktorkorrelationen EFA mit drei Faktoren (Maximum Likelihood, Promax) – Pilottest*

Faktor	1	2	3
1	-		
2	.20	-	
3	.02	.35	-

*Anmerkungen.*

EFA = Explorative Faktorenanalyse.

<sup>1</sup> Dichotome Antwortskala.

<sup>2</sup> Sechsstufige Antwortskala.

Tabelle C.5.

*Reliabilitäten für Alter, Körpergröße und Körpergewicht – Pilottest*

Merkmal	$\alpha$	$\omega$
Alter	.81	.80
Körpergröße	.94	.95
Körpergewicht	.91	.92

*Anmerkung.*

$\alpha$  = Cronbachs Alpha;  $\omega$  = McDonalds Omega.

Tabelle C.6.

*Items für die Merkmale Alter, Körpergröße und Körpergewicht – Haupterhebung*

Merkmal und Item-Nummer		Item
Alter		
1	(D, P)	Ich habe körperliche Gebrechen.
2	(D)	Im Laufe der Zeit hat meine geistige Leistungsfähigkeit abgenommen.
3	(D, P)	Ich habe viel Lebenserfahrung.
4	(D, P)	Im Laufe der Zeit hat meine Reaktionsfähigkeit abgenommen.
5	(D, P)	Im Laufe der Zeit hat meine Merkfähigkeit abgenommen.
6	(D, P)	Ich mache mir viele Gedanken über das Älterwerden.
7	(D)	Den größten Teil meines Lebens habe ich bereits gelebt.
8	(D, P)	Ich habe zunehmend Schwierigkeiten technischen Weiterentwicklungen zu folgen.
9	(D)	Ich habe jahrelange Berufserfahrung.
10	(D, P)*	Mein ganzes Leben liegt noch vor mir.
11	(D, P)	Früher war ich risikofreudiger.
12	(P)	Ich verrichte Alltagstätigkeiten langsamer als früher.
Körpergröße		
13	(D, P)	Ich ziehe den Kopf ein, wenn ich durch Türen gehe.
14	(D)	Inmitten einer Menschenmenge kann ich über die meisten anderen Menschen hinwegblicken.
15	(D)	Stühle und Tische sind meistens zu niedrig für mich.
16	(D, P)	Im Supermarkt komme ich ohne Probleme an die Sachen im obersten Regal.
17	(D, P)	Wenn ich ein Auto fahre, das vorher jemand anderes gefahren hat, muss ich üblicherweise den Sitz nach hinten schieben.
18	(D)	Wenn ich mich in Hotelbetten gerade ausstrecke, dann stehen meine Füße über.
19	(D, P)	Im Flugzeug stoße ich mit den Knien am Vordersitz an.
20	(D, P)	Hosen sind mir häufig zu kurz.
21	(D, P)*	Ich brauche einen Stuhl, wenn ich Sachen aus dem obersten Regal holen möchte.
22	(D, P)*	Wenn ich mich mit anderen Menschen unterhalte, muss ich nach oben schauen.
23	(D, P)*	Ich muss mich bei Gruppenfotos nach vorne stellen, damit ich gut zu sehen bin.
24	(P)	Wenn ich andere Menschen zur Begrüßung umarme, muss ich mich nach unten beugen.
Körpergewicht		
25	(D)	Ich habe oft Angst, dass Stühle unter mir nachgeben.
26	(D)	Möglicherweise denken andere Menschen von mir, dass ich abends auf dem Sofa sitze und Schokolade und Chips esse.
27	(D, P)	Bei Hosen habe ich eine große Bundweite.
28	(D)	Ich bin dick.
29	(D)	Ich wiege viel.
30	(D, P)	Ich müsste abnehmen.

Tabelle C.6. (Fortsetzung)

*Items für die Merkmale Alter, Körpergröße und Körpergewicht – Haupterhebung*

Merkmal und Item-Nummer		Item
31	(D, P)*	Meine Hosen haben eine kleine Bundweite.
32	(D)*	Ich kann mich gut durch enge Spalten quetschen.
33	(D)*	Zwischen den Armlehnen in Flugzeugsitzen habe ich viel Platz.
34	(D)	Im Aufzug nehme ich viel Platz ein.
35	(P)	Wenn ich die Wahl habe zwischen Aufzug und Treppe, dann nehme ich den Aufzug.
36	(P)	Ich sollte weniger essen.

*Anmerkungen.*



D = Items hervorgegangen aus dem deduktiven Testkonstruktionsansatz.

P = Items hervorgegangen aus dem Prototypenansatz.

\* Items mit negativer Polung.

Abbildung C.6. Fragebogen Haupterhebung – Informationsseite

**Informationen zur Befragung**



Sehr geehrte Teilnehmerin, sehr geehrter Teilnehmer,

bitte lesen Sie die folgenden Informationen zur Befragung und zum Schutz der erhobenen Daten durch.

Die Befragung dauert ca. **4 Minuten**.

**1. Ziel der Befragung**

Ziel der Befragung ist die grundlagenwissenschaftliche Untersuchung psychologischer Messinstrumente im Rahmen des Promotionsstudiums der Versuchsleiterin am Lehrstuhl für psychologische Methodenlehre und Diagnostik an der Ludwig-Maximilians-Universität München. Die Teilnahme ist freiwillig und ohne Vergütung. Die Befragung kann jederzeit ohne Nennung von Gründen und ohne Konsequenzen abgebrochen werden.

**2. Daten**

Alle erhobenen Daten werden anonym behandelt. Ihr Name wird an keiner Stelle erfasst. Rückschlüsse auf Ihre Person sind nicht möglich und nicht von Interesse. Die aufgezeichneten Daten lassen sich in folgende Kategorien unterteilen:

- a) Antworten auf Fragebogenitems
- b) Demographische Daten

**3. Aufbewahrung und Löschung der Daten**

Die anonymisierten Daten (Antworten auf Fragebogenitems, Demographische Daten) werden zur Verarbeitung und Auswertung gespeichert. Es werden keine Daten gespeichert, die Rückschlüsse auf die Person zulassen.


Die Ergebnisse und Daten dieser Studie werden im Rahmen des Promotionsstudiums und zur Erstellung der Dissertation der Versuchsleiterin verwendet und anschließend ggf. als wissenschaftliche Publikation veröffentlicht. Dies geschieht in anonymisierter Form, d. h. ohne, dass die Daten einer spezifischen Person zugeordnet werden können. Die vollständig anonymisierten Daten dieser Studie werden als "open data" in einem internetbasierten Repositorium namens Open Science Framework (osf.io) zugänglich gemacht. Damit folgt diese Studie den Empfehlungen der Deutschen Forschungsgemeinschaft (DFG) und der Deutschen Gesellschaft für Psychologie (DGPs) zur Qualitätssicherung in Bezug auf Nachprüfbarkeit und Reproduzierbarkeit wissenschaftlicher Ergebnisse sowie der optimalen Datennachnutzung.

**Wenn Sie mit den oben genannten Punkten zur Datenerhebung einverstanden sind, geht es nun weiter zur Befragung. Sollten Sie mit den oben genannten Punkten nicht einverstanden sein, können Sie die Befragung an dieser Stelle beenden.**

Bei Fragen zur Studie können Sie gerne eine E-Mail schreiben an:  
Kathryn.Eichhorn@psy.lmu.de



Abbildung C.7. Fragebogen Haupterhebung Version 2PS – erste Seite

	<p>LUDWIG- MAXIMILIANS- UNIVERSITÄT MÜNCHEN</p>	<p>Nachfolgend werden Ihnen 36 Aussagen vorgelegt. Markieren Sie bei jeder Aussage, inwiefern Sie dieser zustimmen (<b>0 = trifft nicht zu; 1 = trifft zu</b>). Abschließend werden demografische Daten erfasst.</p> <p>Der Fragebogen hat <b>2 SEITEN!</b> Es ist wichtig den Fragebogen <b>VOLLSTÄNDIG</b> auszufüllen, da nur vollständige Fragebögen für die anschließenden Analysen verwendet werden können. <b>Vielen Dank! :-)</b></p>		
Ich habe viel Lebenserfahrung.	Trifft nicht zu	<input type="checkbox"/> 0	<input type="checkbox"/> 1	Trifft zu
Ich ziehe den Kopf ein, wenn ich durch Türen gehe.	Trifft nicht zu	<input type="checkbox"/> 0	<input type="checkbox"/> 1	Trifft zu
Zwischen den Armlehnen in Flugzeugsitzen habe ich viel Platz.	Trifft nicht zu	<input type="checkbox"/> 0	<input type="checkbox"/> 1	Trifft zu
Ich habe jahrelange Berufserfahrung.	Trifft nicht zu	<input type="checkbox"/> 0	<input type="checkbox"/> 1	Trifft zu
Inmitten einer Menschenmenge kann ich über die meisten anderen Menschen hinwegblicken.	Trifft nicht zu	<input type="checkbox"/> 0	<input type="checkbox"/> 1	Trifft zu
Bei Hosen habe ich eine große Bundweite.	Trifft nicht zu	<input type="checkbox"/> 0	<input type="checkbox"/> 1	Trifft zu
Im Laufe der Zeit hat meine geistige Leistungsfähigkeit abgenommen.	Trifft nicht zu	<input type="checkbox"/> 0	<input type="checkbox"/> 1	Trifft zu
Wenn ich ein Auto fahre, das vorher jemand anderes gefahren hat, muss ich üblicherweise den Sitz nach hinten schieben.	Trifft nicht zu	<input type="checkbox"/> 0	<input type="checkbox"/> 1	Trifft zu
Ich sollte weniger essen.	Trifft nicht zu	<input type="checkbox"/> 0	<input type="checkbox"/> 1	Trifft zu
Mein ganzes Leben liegt noch vor mir.	Trifft nicht zu	<input type="checkbox"/> 0	<input type="checkbox"/> 1	Trifft zu
Im Supermarkt komme ich ohne Probleme an die Sachen im obersten Regal.	Trifft nicht zu	<input type="checkbox"/> 0	<input type="checkbox"/> 1	Trifft zu
Im Aufzug nehme ich viel Platz ein.	Trifft nicht zu	<input type="checkbox"/> 0	<input type="checkbox"/> 1	Trifft zu
Im Laufe der Zeit hat meine Reaktionsfähigkeit abgenommen.	Trifft nicht zu	<input type="checkbox"/> 0	<input type="checkbox"/> 1	Trifft zu
Stühle und Tische sind meistens zu niedrig für mich.	Trifft nicht zu	<input type="checkbox"/> 0	<input type="checkbox"/> 1	Trifft zu
Ich wiege viel.	Trifft nicht zu	<input type="checkbox"/> 0	<input type="checkbox"/> 1	Trifft zu
Ich verrichte Alltagstätigkeiten langsamer als früher.	Trifft nicht zu	<input type="checkbox"/> 0	<input type="checkbox"/> 1	Trifft zu
Wenn ich mich in Hotelbetten gerade ausstrecke, dann stehen meine Füße über.	Trifft nicht zu	<input type="checkbox"/> 0	<input type="checkbox"/> 1	Trifft zu
Meine Hosen haben eine kleine Bundweite.	Trifft nicht zu	<input type="checkbox"/> 0	<input type="checkbox"/> 1	Trifft zu
Im Laufe der Zeit hat meine Merkfähigkeit abgenommen.	Trifft nicht zu	<input type="checkbox"/> 0	<input type="checkbox"/> 1	Trifft zu
Im Flugzeug stoße ich mit den Knien am Vordersitz an.	Trifft nicht zu	<input type="checkbox"/> 0	<input type="checkbox"/> 1	Trifft zu

**Bitte wenden!**

Abbildung C.8. Fragebogen Haupterhebung Version 2PS – zweite Seite

Ich kann mich gut durch enge Spalten quetschen.	Trifft nicht zu	<input type="checkbox"/> 0	<input type="checkbox"/> 1	Trifft zu
Ich habe zunehmend Schwierigkeiten technischen Weiterentwicklungen zu folgen.	Trifft nicht zu	<input type="checkbox"/> 0	<input type="checkbox"/> 1	Trifft zu
Wenn ich mich mit anderen Menschen unterhalte, muss ich nach oben schauen.	Trifft nicht zu	<input type="checkbox"/> 0	<input type="checkbox"/> 1	Trifft zu
Ich habe oft Angst, dass Stühle unter mir nachgeben.	Trifft nicht zu	<input type="checkbox"/> 0	<input type="checkbox"/> 1	Trifft zu
Früher war ich risikofreudiger.	Trifft nicht zu	<input type="checkbox"/> 0	<input type="checkbox"/> 1	Trifft zu
Hosen sind mir häufig zu kurz.	Trifft nicht zu	<input type="checkbox"/> 0	<input type="checkbox"/> 1	Trifft zu
Ich bin dick.	Trifft nicht zu	<input type="checkbox"/> 0	<input type="checkbox"/> 1	Trifft zu
Ich mache mir viele Gedanken über das Älterwerden.	Trifft nicht zu	<input type="checkbox"/> 0	<input type="checkbox"/> 1	Trifft zu
Ich brauche einen Stuhl, wenn ich Sachen aus dem obersten Regal holen möchte.	Trifft nicht zu	<input type="checkbox"/> 0	<input type="checkbox"/> 1	Trifft zu
Wenn ich die Wahl habe zwischen Aufzug und Treppe, dann nehme ich den Aufzug.	Trifft nicht zu	<input type="checkbox"/> 0	<input type="checkbox"/> 1	Trifft zu
Den größten Teil meines Lebens habe ich bereits gelebt.	Trifft nicht zu	<input type="checkbox"/> 0	<input type="checkbox"/> 1	Trifft zu
Ich muss mich bei Gruppenfotos nach vorne stellen, damit ich gut zu sehen bin.	Trifft nicht zu	<input type="checkbox"/> 0	<input type="checkbox"/> 1	Trifft zu
Ich müsste abnehmen.	Trifft nicht zu	<input type="checkbox"/> 0	<input type="checkbox"/> 1	Trifft zu
Ich habe körperliche Gebrechen.	Trifft nicht zu	<input type="checkbox"/> 0	<input type="checkbox"/> 1	Trifft zu
Wenn ich andere Menschen zur Begrüßung umarme, muss ich mich nach unten beugen.	Trifft nicht zu	<input type="checkbox"/> 0	<input type="checkbox"/> 1	Trifft zu
Möglicherweise denken andere Menschen von mir, dass ich abends auf dem Sofa sitze und Schokolade und Chips esse.	Trifft nicht zu	<input type="checkbox"/> 0	<input type="checkbox"/> 1	Trifft zu

Geschlecht: weiblich ☐ männlich ☐

Muttersprache Deutsch: ja ☐ nein ☐

Alter in Jahren: \_\_\_\_\_

Körpergröße  
in Zentimeter: \_\_\_\_\_

Körpergewicht  
in Kilogramm: \_\_\_\_\_

Höchster Bildungsabschluss:

☐ kein Schulabschluss

☐ Hauptschulabschluss oder gleichwertig


☐ Realschule oder gleichwertig

☐ Abitur oder Fachabitur

☐ Universitäts- oder Hochschulabschluss

☐ Promotion oder Habilitation

Abbildung C.9. Fragebogen Haupterhebung Version 6PS – erste Seite

	<p>LUDWIG- MAXIMILIANS- UNIVERSITÄT MÜNCHEN</p>	<p>Nachfolgend werden Ihnen 36 Aussagen vorgelegt. Markieren Sie bei jeder Aussage, inwiefern Sie dieser zustimmen (von <b>1 = trifft nicht zu</b> bis <b>6 = trifft zu</b>). Abschließend werden demographische Daten erfasst.</p> <p>Der Fragebogen hat <b>2 SEITEN</b>! Es ist wichtig den Fragebogen <b>VOLLSTÄNDIG</b> auszufüllen, da nur vollständige Fragebögen für die anschließenden Analysen verwendet werden können. <b>Vielen Dank! :-)</b></p>						
Ich habe viel Lebenserfahrung.	Trifft nicht zu	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> 6	Trifft zu
Ich ziehe den Kopf ein, wenn ich durch Türen gehe.	Trifft nicht zu	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> 6	Trifft zu
Zwischen den Armlehnen in Flugzeugsitzen habe ich viel Platz.	Trifft nicht zu	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> 6	Trifft zu
Ich habe jahrelange Berufserfahrung.	Trifft nicht zu	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> 6	Trifft zu
Inmitten einer Menschenmenge kann ich über die meisten anderen Menschen hinwegblicken.	Trifft nicht zu	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> 6	Trifft zu
Bei Hosen habe ich eine große Bundweite.	Trifft nicht zu	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> 6	Trifft zu
Im Laufe der Zeit hat meine geistige Leistungsfähigkeit abgenommen.	Trifft nicht zu	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> 6	Trifft zu
Wenn ich ein Auto fahre, das vorher jemand anderes gefahren hat, muss ich üblicherweise den Sitz nach hinten schieben.	Trifft nicht zu	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> 6	Trifft zu
Ich sollte weniger essen.	Trifft nicht zu	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> 6	Trifft zu
Mein ganzes Leben liegt noch vor mir.	Trifft nicht zu	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> 6	Trifft zu
Im Supermarkt komme ich ohne Probleme an die Sachen im obersten Regal.	Trifft nicht zu	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> 6	Trifft zu
Im Aufzug nehme ich viel Platz ein.	Trifft nicht zu	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> 6	Trifft zu
Im Laufe der Zeit hat meine Reaktionsfähigkeit abgenommen.	Trifft nicht zu	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> 6	Trifft zu
Stühle und Tische sind meistens zu niedrig für mich.	Trifft nicht zu	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> 6	Trifft zu
Ich wiege viel.	Trifft nicht zu	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> 6	Trifft zu
Ich verrichte Alltagstätigkeiten langsamer als früher.	Trifft nicht zu	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> 6	Trifft zu
Wenn ich mich in Hotelbetten gerade ausstrecke, dann stehen meine Füße über.	Trifft nicht zu	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> 6	Trifft zu
Meine Hosen haben eine kleine Bundweite.	Trifft nicht zu	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> 6	Trifft zu
Im Laufe der Zeit hat meine Merkfähigkeit abgenommen.	Trifft nicht zu	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> 6	Trifft zu
Im Flugzeug stoße ich mit den Knien am Vordersitz an.	Trifft nicht zu	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> 6	Trifft zu

**Bitte wenden!**

Abbildung C.10. Fragebogen Haupterhebung Version 6PS – zweite Seite

Ich kann mich gut durch enge Spalten quetschen.	Trifft nicht zu	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Trifft zu
Ich habe zunehmend Schwierigkeiten technischen Weiterentwicklungen zu folgen.	Trifft nicht zu	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Trifft zu
Wenn ich mich mit anderen Menschen unterhalte, muss ich nach oben schauen.	Trifft nicht zu	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Trifft zu
Ich habe oft Angst, dass Stühle unter mir nachgeben.	Trifft nicht zu	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Trifft zu
Früher war ich risikofreudiger.	Trifft nicht zu	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Trifft zu
Hosen sind mir häufig zu kurz.	Trifft nicht zu	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Trifft zu
Ich bin dick.	Trifft nicht zu	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Trifft zu
Ich mache mir viele Gedanken über das Älterwerden.	Trifft nicht zu	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Trifft zu
Ich brauche einen Stuhl, wenn ich Sachen aus dem obersten Regal holen möchte.	Trifft nicht zu	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Trifft zu
Wenn ich die Wahl habe zwischen Aufzug und Treppe, dann nehme ich den Aufzug.	Trifft nicht zu	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Trifft zu
Den größten Teil meines Lebens habe ich bereits gelebt.	Trifft nicht zu	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Trifft zu
Ich muss mich bei Gruppenfotos nach vorne stellen, damit ich gut zu sehen bin.	Trifft nicht zu	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Trifft zu
Ich müsste abnehmen.	Trifft nicht zu	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Trifft zu
Ich habe körperliche Gebrechen.	Trifft nicht zu	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Trifft zu
Wenn ich andere Menschen zur Begrüßung umarme, muss ich mich nach unten beugen.	Trifft nicht zu	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Trifft zu
Möglicherweise denken andere Menschen von mir, dass ich abends auf dem Sofa sitze und Schokolade und Chips esse.	Trifft nicht zu	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Trifft zu

Geschlecht: weiblich ☐ männlich ☐

Muttersprache Deutsch: ja ☐ nein ☐

Alter in Jahren: \_\_\_\_\_

Körpergröße  
in Zentimeter: \_\_\_\_\_

Körpergewicht  
in Kilogramm: \_\_\_\_\_

Höchster Bildungsabschluss:

☐ kein Schulabschluss

☐ Hauptschulabschluss oder gleichwertig

☐ Realschule oder gleichwertig

☐ Abitur oder Fachabitur

☐ Universitäts- oder Hochschulabschluss

☐ Promotion oder Habilitation

Abbildung C.11. Histogramme der realen Werte für das Merkmal Alter

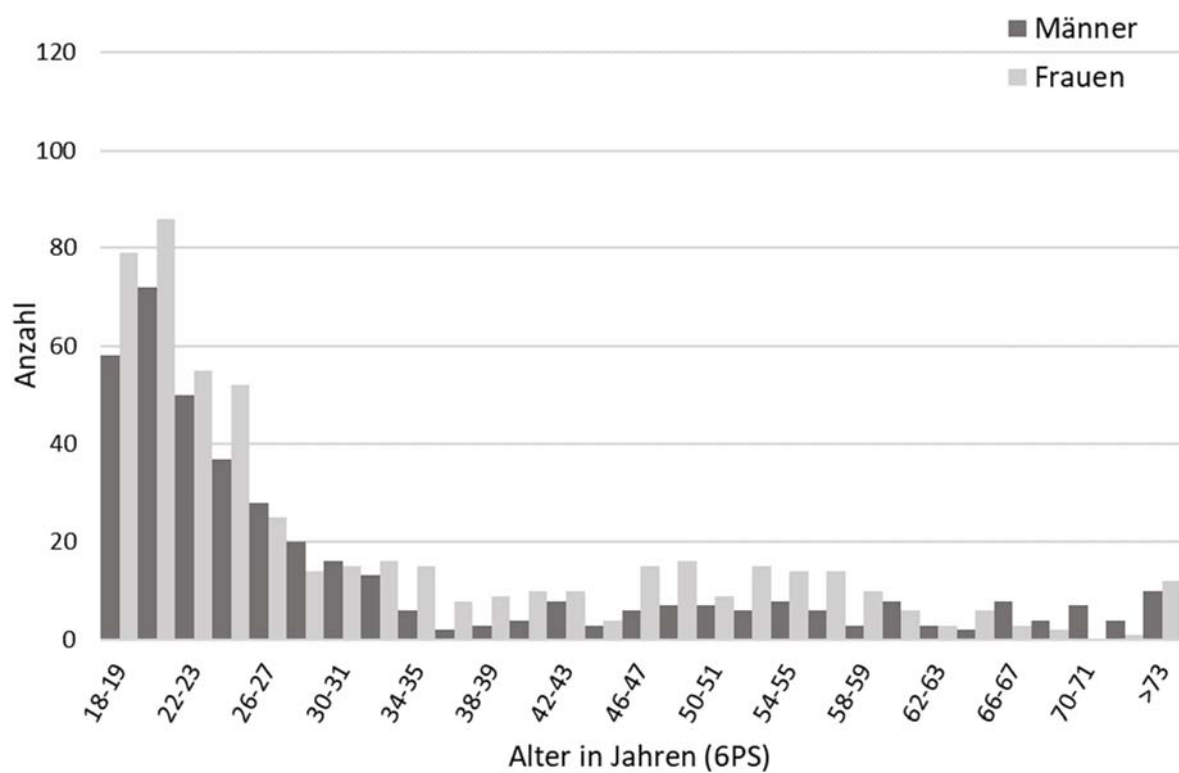
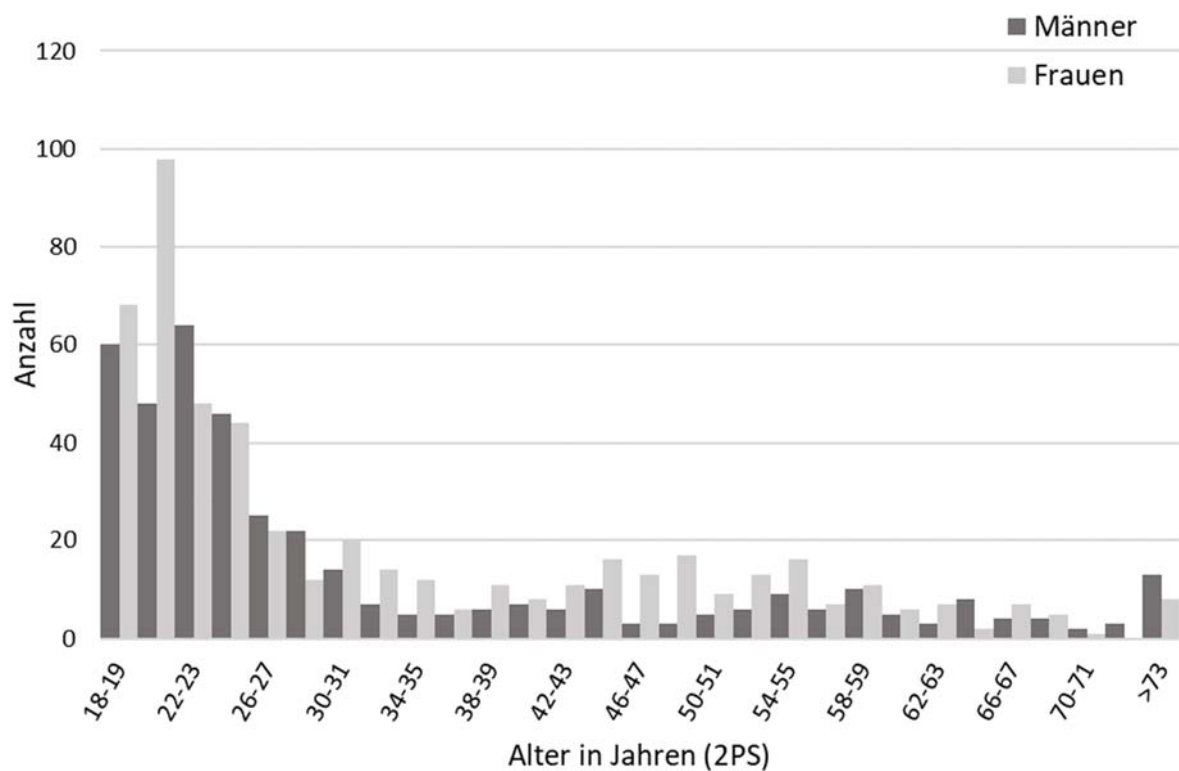


Abbildung C.12. Histogramme der realen Werte für das Merkmal Körpergröße

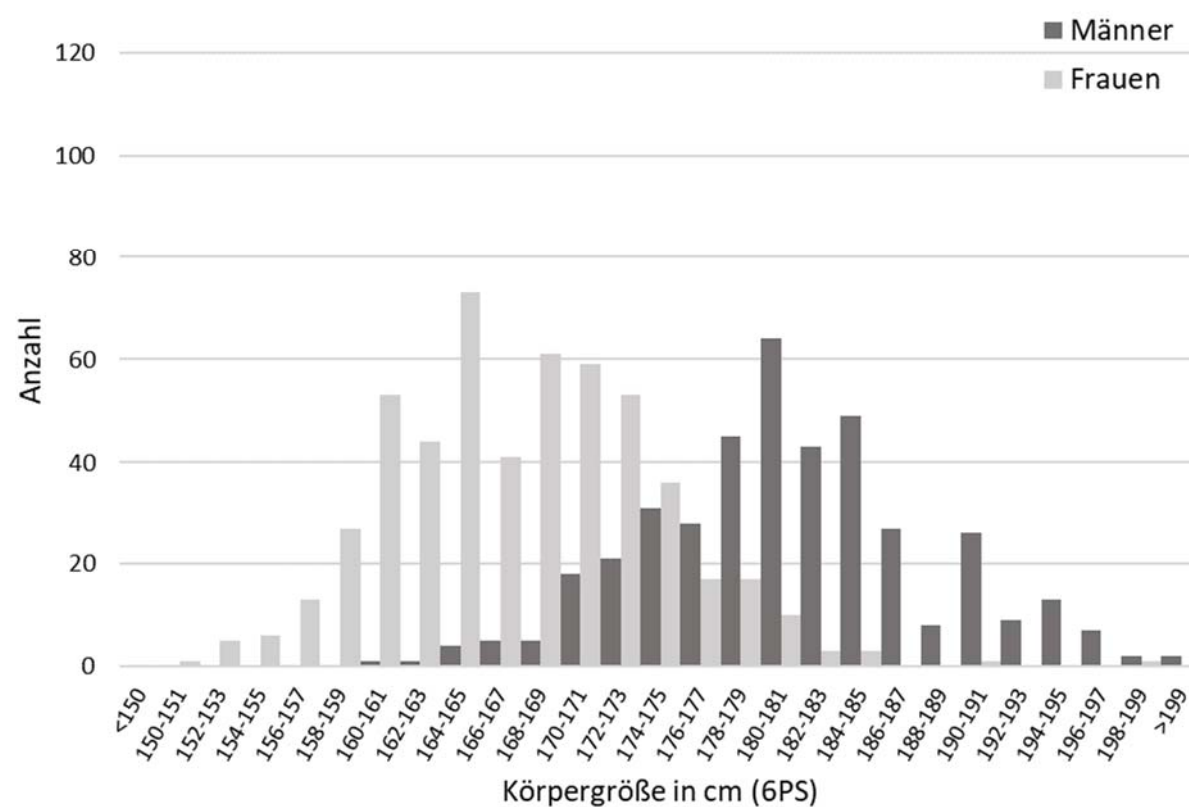
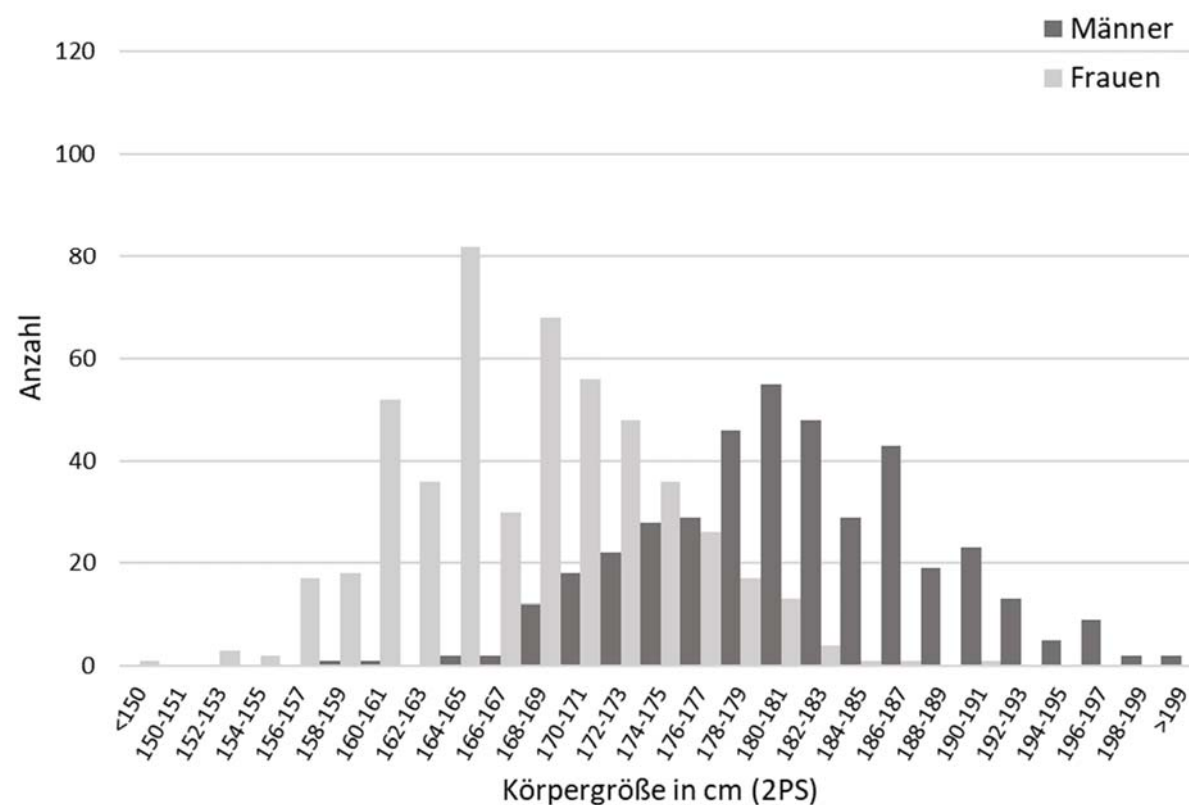


Abbildung C.13. Histogramme der realen Werte für das Merkmal Körpergewicht

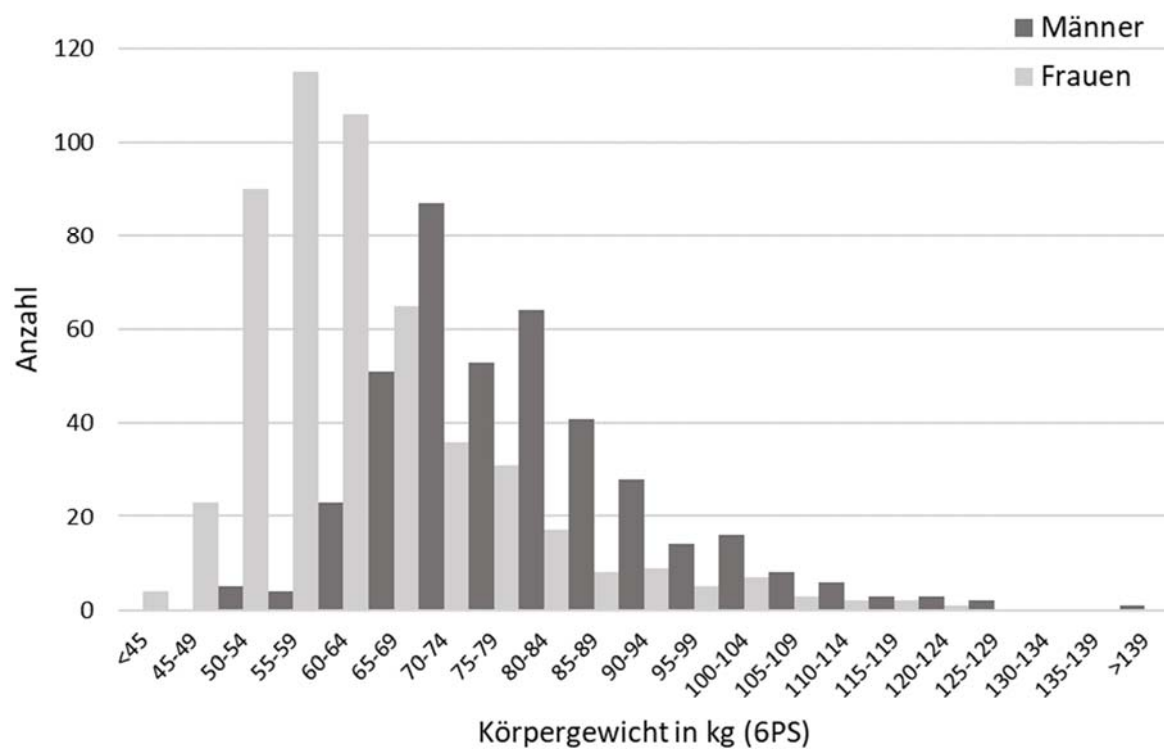
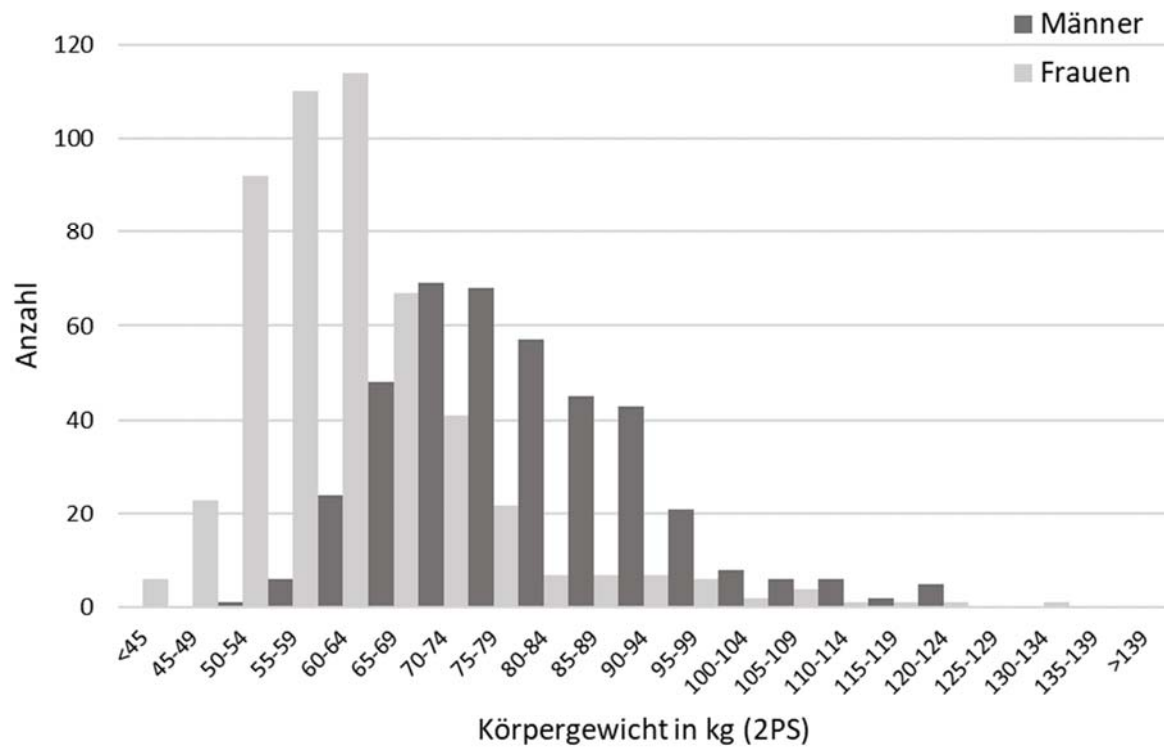


Tabelle C.7.

Punktbiseriale Korrelationen ( $r_{pb}$ ) Alter, Körpergröße, Körpergewicht und alle Items – Männer, Bedingung 2PS<sup>1</sup>

Merkmal und Item-Nummer	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1. Alter	-																		
2. Größe	-.15	-																	
3. Gewicht	.23	.50	-																
4. A_1	.25	-.005	.16	-															
5. A_2	.18	-.09	-.13	.18	-														
6. A_3	.45	-.08	.22	.09	-.003	-													
7. A_4	.44	-.11	.09	.23	.37	.12	-												
8. A_5	.35	-.11	.04	.24	.51	.14	.43	-											
9. A_6	.01	.03	.02	.14	.03	.10	-.001	.10	-										
10. A_7	.77	-.17	.11	.22	.20	.33	.38	.33	.05	-									
11. A_8	.43	-.10	.10	.20	.17	.21	.25	.22	.10	.44	-								
12. A_9	.69	-.16	.19	.08	.12	.47	.27	.25	.02	.52	.28	-							
13. A_10	.56	-.07	.17	.19	.20	.23	.33	.31	-.007	.55	.32	.43	-						
14. A_11	.25	-.04	.05	.10	.14	.18	.23	.24	.12	.24	.20	.22	.23	-					
15. A_12	.37	-.19	.03	.22	.37	.15	.41	.37	.04	.36	.22	.20	.28	.20	-				
16. Kgr_13	-.09	.50	.32	.02	.01	-.06	-.001	-.08	-.02	-.09	.009	-.08	-.08	-.09	-.10	-			
17. Kgr_14	-.09	.60	.34	-.06	-.05	-.02	-.10	-.08	-.002	-.15	-.12	-.06	-.10	.008	-.06	.35	-		
18. Kgr_15	-.05	.50	.24	-.01	-.04	-.02	.006	-.08	.01	-.10	-.08	-.04	-.10	.01	-.08	.43	.41	-	
19. Kgr_16	-.06	.43	.23	-.05	-.04	-.01	-.05	-.03	.04	-.01	-.07	-.04	.02	-.09	-.02	.10	.24	.10	-
20. Kgr_17	.008	.50	.25	-.05	-.01	-.04	-.04	-.04	.06	-.001	-.005	-.03	.02	.01	.006	.29	.43	.24	.24
21. Kgr_18	-.09	.58	.30	-.04	-.06	-.05	-.05	-.11	-.04	-.12	-.09	-.11	-.13	-.04	-.15	.49	.48	.45	.15
22. Kgr_19	-.03	.60	.34	.05	-.06	-.04	-.04	-.002	.04	-.10	-.06	-.09	-.005	.001	-.04	.30	.45	.40	.20
23. Kgr_20	-.12	.41	.07	-.03	-.08	-.12	-.07	-.03	-.03	-.09	-.09	-.08	-.07	-.08	-.15	.28	.34	.40	.03
24. Kgr_21	-.17	.38	.23	-.08	-.13	-.03	-.12	-.12	.05	-.16	-.13	-.15	-.09	-.09	-.11	.13	.26	.13	.43
25. Kgr_22	.007	.39	.23	.05	-.02	.07	.03	-.09	.05	.01	-.02	-.007	.03	-.04	-.04	.15	.19	.13	.41
26. Kgr_23	.04	.46	.30	.04	-.01	.000	-.06	-.01	.06	-.02	-.08	-.02	.04	-.04	-.02	.16	.31	.16	.38
27. Kgr_24	-.13	.62	.29	-.02	.002	-.09	-.009	-.05	.06	-.16	-.11	-.11	-.08	-.03	-.09	.40	.54	.45	.24
28. Kge_25	.09	.10	.25	.13	-.004	.06	.15	.07	.06	.03	.04	.06	.08	-.01	-.006	.19	.05	.17	-.05
29. Kge_26	-.06	.05	.22	.08	.05	-.03	.05	.05	.05	-.10	.04	-.003	.02	.09	.08	.03	.08	.07	-.09
30. Kge_27	.08	.13	.47	.12	.08	.05	.11	.08	.04	.11	.18	.02	.04	.10	.16	.20	.13	.10	.01
31. Kge_28	.20	.007	.40	.19	.13	.08	.15	.12	.02	.16	.20	.13	.16	.08	.18	.07	.03	.02	-.02
32. Kge_29	.17	.12	.57	.24	.10	.14	.13	.12	.08	.12	.18	.15	.10	.12	.12	.17	.07	.08	.02
33. Kge_30	.21	.02	.51	.24	.11	.19	.14	.10	.10	.11	.15	.14	.16	.10	.17	.007	-.02	-.01	.02
34. Kge_31	.08	.12	.43	.08	.06	.15	.05	.14	.06	.03	.11	.006	.07	.08	.05	.08	.04	-.009	.09
35. Kge_32	.08	.13	.39	.10	.11	.07	.08	.03	.02	.06	.14	.08	.12	-.001	.08	.17	.06	.05	-.02
36. Kge_33	-.03	.24	.30	.06	.02	.06	-.004	.007	.08	-.03	.009	.05	.02	.07	.02	.21	.22	.14	.004
37. Kge_34	.02	.19	.30	.11	.06	.07	.09	.04	.006	-.02	.06	.03	.02	-.04	.04	.15	.18	.18	.01
38. Kge_35	.10	-.002	.16	.04	.11	.09	.12	.09	.07	.09	.12	.05	.11	.03	.12	.04	-.02	-.03	.09
39. Kge_36	.18	-.01	.40	.17	.12	.17	.06	.08	.06	.12	.13	.18	.15	.14	.13	.03	.009	.01	.02



Tabelle C.7. (Fortsetzung)

Punktbiserial Korrelationen ( $r_{pb}$ ) Alter, Körpergröße, Körpergewicht und alle Items – Männer, Bedingung 2PS<sup>1</sup>

Merkmal und Item-Nummer	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39
1. Alter																				
2. Größe																				
3. Gewicht																				
4. A_1																				
5. A_2																				
6. A_3																				
7. A_4																				
8. A_5																				
9. A_6																				
10. A_7																				
11. A_8																				
12. A_9																				
13. A_10																				
14. A_11																				
15. A_12																				
16. Kgr_13																				
17. Kgr_14																				
18. Kgr_15																				
19. Kgr_16																				
20. Kgr_17	-																			
21. Kgr_18	<b>.38</b>	-																		
22. Kgr_19	<b>.41</b>	<b>.50</b>	-																	
23. Kgr_20	<b>.25</b>	<b>.45</b>	<b>.30</b>	-																
24. Kgr_21	<b>.21</b>	<b>.17</b>	<b>.26</b>	<b>.08</b>	-															
25. Kgr_22	<b>.23</b>	<b>.12</b>	<b>.23</b>	<b>.04</b>	<b>.39</b>	-														
26. Kgr_23	<b>.28</b>	<b>.22</b>	<b>.34</b>	<b>.09</b>	<b>.44</b>	<b>.53</b>	-													
27. Kgr_24	<b>.41</b>	<b>.54</b>	<b>.50</b>	<b>.40</b>	<b>.23</b>	<b>.16</b>	<b>.28</b>	-												
28. Kge_25	<b>.12</b>	<b>.15</b>	<b>.13</b>	<b>.09</b>	<b>.01</b>	<b>-.02</b>	<b>.01</b>	<b>.10</b>	-											
29. Kge_26	<b>-.004</b>	<b>-.004</b>	<b>.05</b>	<b>.01</b>	<b>-.05</b>	<b>-.02</b>	<b>.03</b>	<b>.09</b>	<b>.09</b>	-										
30. Kge_27	<b>.11</b>	<b>.11</b>	<b>.15</b>	<b>.01</b>	<b>.08</b>	<b>.04</b>	<b>.08</b>	<b>.12</b>	<b>.21</b>	<b>.26</b>	-									
31. Kge_28	<b>.05</b>	<b>.01</b>	<b>.02</b>	<b>-.02</b>	<b>-.03</b>	<b>-.009</b>	<b>.08</b>	<b>.04</b>	<b>.08</b>	<b>.34</b>	<b>.40</b>	-								
32. Kge_29	<b>.09</b>	<b>.08</b>	<b>.11</b>	<b>.008</b>	<b>.05</b>	<b>-.001</b>	<b>.12</b>	<b>.09</b>	<b>.26</b>	<b>.32</b>	<b>.50</b>	<b>.58</b>	-							
33. Kge_30	<b>-.02</b>	<b>-.03</b>	<b>.03</b>	<b>-.11</b>	<b>.03</b>	<b>.02</b>	<b>.09</b>	<b>.02</b>	<b>.13</b>	<b>.27</b>	<b>.39</b>	<b>.50</b>	<b>.59</b>	-						
34. Kge_31	<b>.08</b>	<b>.02</b>	<b>.11</b>	<b>-.08</b>	<b>.09</b>	<b>.11</b>	<b>.23</b>	<b>.06</b>	<b>.09</b>	<b>.15</b>	<b>.39</b>	<b>.24</b>	<b>.36</b>	<b>.42</b>	-					
35. Kge_32	<b>.10</b>	<b>.06</b>	<b>.14</b>	<b>.07</b>	<b>.08</b>	<b>.08</b>	<b>.12</b>	<b>.13</b>	<b>.10</b>	<b>.13</b>	<b>.32</b>	<b>.19</b>	<b>.38</b>	<b>.30</b>	<b>.42</b>	-				
36. Kge_33	<b>.13</b>	<b>.17</b>	<b>.24</b>	<b>.08</b>	<b>.09</b>	<b>.09</b>	<b>.07</b>	<b>.16</b>	<b>.08</b>	<b>.02</b>	<b>.19</b>	<b>.05</b>	<b>.17</b>	<b>.12</b>	<b>.15</b>	<b>.22</b>	-			
37. Kge_34	<b>.14</b>	<b>.17</b>	<b>.17</b>	<b>.16</b>	<b>.05</b>	<b>.000</b>	<b>.03</b>	<b>.20</b>	<b>.19</b>	<b>.11</b>	<b>.35</b>	<b>.24</b>	<b>.23</b>	<b>.19</b>	<b>.17</b>	<b>.17</b>	<b>.13</b>	-		
38. Kge_35	<b>-.009</b>	<b>-.02</b>	<b>.005</b>	<b>-.06</b>	<b>-.01</b>	<b>.07</b>	<b>.002</b>	<b>.01</b>	<b>.08</b>	<b>.08</b>	<b>.09</b>	<b>.16</b>	<b>.13</b>	<b>.23</b>	<b>.07</b>	<b>.09</b>	<b>.06</b>	<b>.09</b>	-	
39. Kge_36	<b>-.007</b>	<b>-.02</b>	<b>.02</b>	<b>-.10</b>	<b>.04</b>	<b>.02</b>	<b>.09</b>	<b>-.01</b>	<b>.15</b>	<b>.24</b>	<b>.27</b>	<b>.36</b>	<b>.49</b>	<b>.67</b>	<b>.31</b>	<b>.23</b>	<b>.07</b>	<b>.12</b>	<b>.24</b>	-

Anmerkungen.

<sup>1</sup> Dichotome Antwortskala; die Nummerierung der Items entspricht der Nummerierung in Tabelle C.6.

Korrelationen innerhalb derselben Skala sind farblich markiert (Alter = rot; Körpergröße = grün;

Körpergewicht = blau); signifikante Ergebnisse ( $p < .01$ ) sind fett gedruckt.

Tabelle C.8.

Punktbiseriale Korrelationen ( $r_{pb}$ ) Alter, Körpergröße, Körpergewicht und alle Items – Frauen, Bedingung 2PS<sup>1</sup>

Merkmal und Item-Nummer	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1. Alter	-																		
2. Größe	-.08	-																	
3. Gewicht	.24	.37	-																
4. A_1	.20	.08	.21	-															
5. A_2	.26	.003	.08	.16	-														
6. A_3	.54	-.02	.19	.11	.18	-													
7. A_4	.40	.005	.09	.17	.48	.22	-												
8. A_5	.42	-.01	.10	.18	.54	.22	.54	-											
9. A_6	-.02	-.001	-.008	.12	.15	.02	.13	.07	-										
10. A_7	.74	-.07	.17	.16	.31	.40	.38	.37	.09	-									
11. A_8	.29	-.08	.09	.12	.29	.21	.28	.30	.15	.30	-								
12. A_9	.67	.004	.31	.18	.22	.53	.31	.27	-.07	.43	.21	-							
13. A_10	.67	-.04	.21	.20	.26	.35	.37	.36	.01	.57	.29	.49	-						
14. A_11	.29	.02	.08	.13	.17	.13	.18	.22	.11	.22	.20	.24	.29	-					
15. A_12	.40	-.007	.17	.28	.35	.25	.39	.46	.10	.39	.31	.22	.28	.17	-				
16. Kgr_13	-.07	.20	.06	-.08	.11	.03	.06	.04	.04	-.06	-.02	-.03	-.09	.03	-.02	-			
17. Kgr_14	-.07	.46	.13	-.02	-.01	-.006	.05	.03	.03	-.08	-.02	-.03	-.10	.02	-.003	.22	-		
18. Kgr_15	.003	.41	.18	-.006	.05	.03	.08	.12	-.01	-.03	.01	.03	-.05	-.04	.08	.25	.39	-	
19. Kgr_16	.06	.63	.26	.07	.02	.06	.04	.03	.004	.05	.03	.05	.05	.03	.06	.10	.36	.27	-
20. Kgr_17	.14	.42	.24	.02	.09	.08	.13	.09	-.02	.09	.07	.09	.09	.02	.11	.16	.36	.36	.37
21. Kgr_18	-.05	.41	.17	.001	.05	.02	.05	.02	.04	-.07	.02	-.007	-.04	-.05	.009	.23	.35	.39	.23
22. Kgr_19	.001	.51	.23	.09	.01	.05	.03	.04	-.02	-.04	.01	.08	-.04	.05	.05	.17	.37	.45	.34
23. Kgr_20	-.06	.57	.19	.08	-.01	.01	.03	-.05	.05	-.08	-.06	.000	-.007	.01	.02	.16	.41	.37	.37
24. Kgr_21	-.08	.64	.23	.04	.03	-.007	.02	.02	-.01	-.03	-.01	-.03	.01	.003	-.02	.12	.35	.22	.64
25. Kgr_22	.13	.57	.30	.05	.03	.14	.04	.08	-.02	.10	.05	.13	.10	.07	.09	.09	.28	.18	.53
26. Kgr_23	.09	.70	.35	.09	.01	.07	.05	.04	-.01	.07	-.005	.12	.09	.07	.08	.11	.30	.21	.59
27. Kgr_24	-.06	.52	.21	-.005	-.06	.01	-.04	-.06	.03	-.11	-.05	-.03	-.10	-.02	-.02	.18	.46	.46	.33
28. Kge_25	.07	.07	.26	.08	.02	.05	.05	-.02	-.007	-.009	.04	.09	.10	.02	.002	-.02	-.005	.05	.007
29. Kge_26	-.15	-.06	.20	.10	-.02	-.16	-.04	-.07	.05	-.11	.03	-.06	-.07	-.05	-.01	-.09	-.03	-.01	.02
30. Kge_27	.08	.10	.57	.10	.04	.11	.05	.06	-.03	.07	.13	.12	.10	.03	.12	.04	-.001	.03	.07
31. Kge_28	.13	.01	.58	.14	.12	.10	.11	.04	.01	.07	.14	.22	.19	.07	.07	-.004	-.06	.001	.008
32. Kge_29	.17	.04	.62	.15	.10	.12	.13	.08	.04	.13	.14	.19	.18	.06	.14	.002	-.09	.000	.04
33. Kge_30	.13	-.01	.52	.17	.11	.09	.07	.05	.06	.11	.13	.19	.13	.07	.14	.03	-.05	-.05	.02
34. Kge_31	.10	.17	.53	.11	.07	.08	.10	.09	.004	.07	.06	.13	.13	.03	.09	.001	.009	-.01	.15
35. Kge_32	.17	.13	.47	.04	.12	.09	.12	.09	.02	.13	.13	.14	.17	.02	.17	-.03	.000	.09	.08
36. Kge_33	.01	.21	.34	.04	.05	.06	.02	.06	-.04	.03	.02	.05	.12	.003	.08	.09	.14	.13	.15
37. Kge_34	.12	.03	.30	.09	.04	.04	.05	.07	.04	.08	.03	.07	.14	.10	.10	-.03	.009	.09	.02
38. Kge_35	-.07	-.03	.04	.07	-.04	-.06	-.02	.03	.09	-.02	.04	-.12	-.07	.02	.11	.02	-.002	.000	.007
39. Kge_36	.03	-.05	.33	.08	.11	.02	.06	.04	.07	.06	.11	.07	.08	.09	.11	.02	-.02	-.02	-.004

Tabelle C.8. (Fortsetzung)

*Punktbiseriale Korrelationen ( $r_{pb}$ ) Alter, Körpergröße, Körpergewicht und alle Items – Frauen, Bedingung 2PS<sup>1</sup>*

Merkmal und Item-Nummer	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39
1. Alter																				
2. Größe																				
3. Gewicht																				
4. A_1																				
5. A_2																				
6. A_3																				
7. A_4																				
8. A_5																				
9. A_6																				
10. A_7																				
11. A_8																				
12. A_9																				
13. A_10																				
14. A_11																				
15. A_12																				
16. Kgr_13																				
17. Kgr_14																				
18. Kgr_15																				
19. Kgr_16																				
20. Kgr_17	-																			
21. Kgr_18	<b>.37</b>	-																		
22. Kgr_19	<b>.38</b>	<b>.40</b>	-																	
23. Kgr_20	<b>.38</b>	<b>.36</b>	<b>.50</b>	-																
24. Kgr_21	<b>.33</b>	<b>.26</b>	<b>.31</b>	<b>.35</b>	-															
25. Kgr_22	<b>.26</b>	<b>.15</b>	<b>.27</b>	<b>.27</b>	<b>.54</b>	-														
26. Kgr_23	<b>.29</b>	<b>.21</b>	<b>.32</b>	<b>.35</b>	<b>.57</b>	<b>.64</b>	-													
27. Kgr_24	<b>.40</b>	<b>.46</b>	<b>.46</b>	<b>.54</b>	<b>.32</b>	<b>.23</b>	<b>.28</b>	-												
28. Kge_25	.03	<b>.12</b>	.04	.07	.04	.07	.03	.06	-											
29. Kge_26	-.02	.000	.02	.007	-.03	-.09	-.03	.05	<b>.19</b>	-										
30. Kge_27	.11	.10	<b>.14</b>	.04	.04	.07	.10	.01	<b>.15</b>	<b>.23</b>	-									
31. Kge_28	.06	.05	.07	.02	.000	.05	.04	-.006	<b>.26</b>	<b>.27</b>	<b>.52</b>	-								
32. Kge_29	.06	.05	.05	-.01	.05	.10	.09	.01	<b>.22</b>	<b>.27</b>	<b>.52</b>	<b>.71</b>	-							
33. Kge_30	-.01	-.02	-.004	-.08	-.02	.03	.05	-.06	<b>.13</b>	<b>.22</b>	<b>.39</b>	<b>.44</b>	<b>.55</b>	-						
34. Kge_31	.11	.06	.07	-.006	<b>.17</b>	<b>.16</b>	<b>.24</b>	-.008	<b>.09</b>	<b>.15</b>	<b>.48</b>	<b>.31</b>	<b>.40</b>	<b>.41</b>	-					
35. Kge_32	<b>.17</b>	.07	.10	.10	.11	.10	<b>.16</b>	.08	<b>.05</b>	.10	<b>.38</b>	<b>.31</b>	<b>.34</b>	<b>.33</b>	<b>.32</b>	-				
36. Kge_33	<b>.12</b>	<b>.13</b>	<b>.19</b>	<b>.14</b>	.10	<b>.22</b>	<b>.16</b>	<b>.13</b>	<b>.05</b>	<b>.08</b>	<b>.24</b>	<b>.19</b>	<b>.23</b>	<b>.18</b>	<b>.27</b>	<b>.17</b>	-			
37. Kge_34	.06	-.05	.10	.09	.002	-.001	-.001	.05	<b>.21</b>	.11	<b>.34</b>	<b>.34</b>	<b>.26</b>	<b>.15</b>	<b>.19</b>	<b>.23</b>	<b>.11</b>	-		
38. Kge_35	-.004	-.03	.05	.05	-.04	-.05	-.08	-.002	<b>.05</b>	<b>.18</b>	.07	.03	.007	.03	-.05	.11	-.01	.07	-	
39. Kge_36	-.02	-.05	-.01	-.05	-.02	-.01	-.003	-.02	<b>.12</b>	<b>.25</b>	<b>.22</b>	<b>.31</b>	<b>.32</b>	<b>.52</b>	<b>.27</b>	<b>.18</b>	<b>.10</b>	<b>.12</b>	<b>.12</b>	-

*Anmerkungen.*<sup>1</sup> Dichotome Antwortskala; die Nummerierung der Items entspricht der Nummerierung in Tabelle C.6.Korrelationen innerhalb derselben Skala sind farblich markiert (Alter = rot; Körpergröße = grün; Körpergewicht = blau); signifikante Ergebnisse ( $p < .01$ ) sind fett gedruckt.

Tabelle C.9.

*Spearman-Korrelationen ( $r_s$ ) Alter, Körpergröße, Körpergewicht und alle Items – Männer, Bedingung 6PS<sup>1</sup>*

Merkmal und Item-Nummer	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1. Alter	-																		
2. Größe	-.14	-																	
3. Gewicht	.29	.48	-																
4. A_1	.17	.05	.16	-															
5. A_2	.26	-.07	.10	.29	-														
6. A_3	.46	-.07	.13	.05	.15	-													
7. A_4	.30	-.01	.22	.29	.57	.09	-												
8. A_5	.28	-.06	.12	.23	.67	.12	.61	-											
9. A_6	.07	-.06	.02	.15	.17	.01	.20	.15	-										
10. A_7	.60	-.11	.16	.31	.38	.39	.39	.41	.09	-									
11. A_8	.31	-.10	.11	.24	.37	.17	.38	.34	.16	.42	-								
12. A_9	.73	-.10	.25	.16	.25	.52	.20	.24	.09	.56	.31	-							
13. A_10	.60	-.11	.13	.19	.23	.30	.28	.28	-.005	.61	.27	.47	-						
14. A_11	.31	-.07	.17	.24	.34	.11	.38	.32	.11	.38	.29	.25	.25	-					
15. A_12	.27	-.09	.13	.30	.52	.16	.52	.52	.19	.43	.40	.25	.25	.41	-				
16. Kgr_13	-.22	.48	.16	.09	.04	-.08	.02	.03	.03	-.04	.03	-.14	-.04	-.02	-.004	-			
17. Kgr_14	-.15	.70	.31	-.01	-.08	.02	-.08	-.06	-.09	-.07	-.08	-.07	-.11	-.07	-.07	.46	-		
18. Kgr_15	-.19	.48	.18	.18	.08	-.12	.10	.09	.02	-.06	.02	-.13	-.12	-.01	.11	.43	.43	-	
19. Kgr_16	-.06	.53	.27	-.03	-.005	.05	-.07	-.05	-.06	-.06	-.08	.003	-.06	.02	-.07	.26	.49	.24	-
20. Kgr_17	.06	.58	.45	-.002	-.01	-.02	.06	.004	-.07	.04	-.06	.05	-.006	-.02	.02	.30	.44	.39	.40
21. Kgr_18	-.18	.63	.34	.14	.003	-.11	.12	.04	-.04	-.04	-.04	-.14	-.11	.02	.04	.47	.51	.56	.32
22. Kgr_19	-.08	.59	.31	.07	-.008	-.06	.08	.006	-.09	.003	-.11	-.07	-.04	.002	.000	.37	.50	.47	.37
23. Kgr_20	-.12	.49	.16	.22	.08	-.07	.07	.07	.002	.02	.08	-.08	-.008	.03	.10	.41	.41	.49	.24
24. Kgr_21	-.15	.48	.33	-.02	-.13	-.02	-.08	-.13	-.09	-.20	-.14	-.08	-.17	-.10	-.11	.11	.41	.19	.56
25. Kgr_22	.03	.52	.37	-.08	-.08	.003	-.05	-.04	-.10	-.08	-.07	.01	-.04	-.06	-.06	.15	.47	.13	.39
26. Kgr_23	.03	.59	.36	-.10	-.08	.03	-.02	-.08	-.06	-.08	-.08	.008	-.007	-.03	-.08	.20	.49	.16	.49
27. Kgr_24	-.22	.66	.25	.09	.02	-.06	.02	.06	-.02	-.04	-.07	-.13	-.06	.03	.003	.45	.54	.46	.42
28. Kge_25	.04	.08	.18	.18	.12	.000	.14	.03	.08	.09	.16	.07	.09	.13	.11	.17	.07	.19	-.14
29. Kge_26	-.12	-.03	.16	.17	.09	-.07	-.08	.04	.08	.05	.06	-.03	.03	.15	.12	.02	-.02	.08	-.08
30. Kge_27	.09	.17	.57	.26	.16	.07	.29	.15	.01	.18	.13	.10	.09	.15	.16	.16	.14	.15	.03
31. Kge_28	.16	.04	.52	.30	.19	.04	.29	.16	.12	.22	.19	.17	.16	.18	.23	.09	-.04	.09	-.09
32. Kge_29	.11	.13	.63	.21	.16	.05	.31	.20	.08	.20	.16	.10	.10	.22	.24	.09	.06	.17	.02
33. Kge_30	.25	-.05	.50	.31	.26	.09	.34	.22	.14	.27	.18	.23	.19	.23	.30	-.004	-.12	.02	-.10
34. Kge_31	.19	.17	.61	.16	.11	.06	.15	.07	-.003	.15	.10	.20	.13	.14	.10	.07	.14	.02	.11
35. Kge_32	.07	.17	.49	.18	.07	.02	.18	.06	.06	.08	.05	.10	.03	.13	.10	.07	.09	.07	.06
36. Kge_33	.04	.24	.36	.08	.02	.002	.18	.05	.12	.05	.07	.06	-.008	.08	.11	.15	.16	.17	.09
37. Kge_34	-.12	.31	.47	.22	.10	-.07	.22	.06	.07	.02	.05	-.03	-.09	.05	.12	.16	.18	.31	.09
38. Kge_35	-.04	.001	.08	.08	.07	-.04	.08	.04	.05	.01	.10	-.05	-.05	.14	.13	.03	-.01	.04	.001
39. Kge_36	.17	-.02	.43	.20	.15	.04	.26	.17	.12	.19	.12	.14	.11	.26	.23	-.02	-.07	.06	-.06

Tabelle C.9. (Fortsetzung)

*Spearman-Korrelationen ( $r_s$ ) Alter, Körpergröße, Körpergewicht und alle Items – Männer, Bedingung 6PS<sup>1</sup>*

Merkmal und Item-Nummer	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39
1. Alter																				
2. Größe																				
3. Gewicht																				
4. A_1																				
5. A_2																				
6. A_3																				
7. A_4																				
8. A_5																				
9. A_6																				
10. A_7																				
11. A_8																				
12. A_9																				
13. A_10																				
14. A_11																				
15. A_12																				
16. Kgr_13																				
17. Kgr_14																				
18. Kgr_15																				
19. Kgr_16																				
20. Kgr_17	-																			
21. Kgr_18	<b>.45</b>	-																		
22. Kgr_19	<b>.49</b>	<b>.56</b>	-																	
23. Kgr_20	<b>.29</b>	<b>.51</b>	<b>.40</b>	-																
24. Kgr_21	<b>.36</b>	<b>.22</b>	<b>.28</b>	<b>.12</b>	-															
25. Kgr_22	<b>.41</b>	<b>.20</b>	<b>.35</b>	<b>.10</b>	<b>.47</b>	-														
26. Kgr_23	<b>.44</b>	<b>.31</b>	<b>.36</b>	<b>.14</b>	<b>.51</b>	<b>.59</b>	-													
27. Kgr_24	<b>.43</b>	<b>.57</b>	<b>.53</b>	<b>.52</b>	<b>.28</b>	<b>.27</b>	<b>.33</b>	-												
28. Kge_25	.10	<b>.18</b>	.07	<b>.18</b>	-.10	<b>-.13</b>	-.10	.08	-											
29. Kge_26	-.04	.12	.002	.07	-.05	<b>-.13</b>	<b>-.14</b>	.09	<b>.21</b>	-										
30. Kge_27	<b>.24</b>	<b>.24</b>	<b>.20</b>	<b>.14</b>	.08	.06	.10	<b>.13</b>	<b>.32</b>	<b>.29</b>	-									
31. Kge_28	.13	<b>.16</b>	.08	<b>.13</b>	-.05	-.03	-.03	.04	<b>.37</b>	<b>.40</b>	<b>.62</b>	-								
32. Kge_29	<b>.20</b>	<b>.22</b>	<b>.15</b>	.10	.05	.05	.08	.13	<b>.28</b>	<b>.33</b>	<b>.65</b>	<b>.68</b>	-							
33. Kge_30	.12	.05	.02	.02	-.04	-.02	-.09	-.04	<b>.26</b>	<b>.37</b>	<b>.56</b>	<b>.78</b>	<b>.66</b>	-						
34. Kge_31	<b>.26</b>	.11	<b>.15</b>	-.004	<b>.19</b>	<b>.17</b>	<b>.15</b>	.06	<b>.19</b>	<b>.23</b>	<b>.63</b>	<b>.51</b>	<b>.57</b>	<b>.50</b>	-					
35. Kge_32	<b>.26</b>	<b>.18</b>	<b>.19</b>	.11	<b>.14</b>	<b>.17</b>	<b>.15</b>	.10	<b>.20</b>	<b>.24</b>	<b>.44</b>	<b>.43</b>	<b>.45</b>	<b>.37</b>	<b>.51</b>	-				
36. Kge_33	<b>.23</b>	<b>.19</b>	<b>.36</b>	.13	.11	<b>.17</b>	<b>.15</b>	<b>.15</b>	<b>.13</b>	<b>.09</b>	<b>.32</b>	<b>.19</b>	<b>.28</b>	<b>.17</b>	<b>.30</b>	<b>.35</b>	-			
37. Kge_34	<b>.31</b>	<b>.38</b>	<b>.31</b>	<b>.22</b>	<b>.19</b>	.05	.008	<b>.28</b>	<b>.30</b>	<b>.28</b>	<b>.48</b>	<b>.41</b>	<b>.49</b>	<b>.35</b>	<b>.37</b>	<b>.41</b>	<b>.34</b>	-		
38. Kge_35	.12	.04	-.03	-.004	-.02	-.07	-.01	-.004	<b>.15</b>	<b>.19</b>	<b>.13</b>	<b>.15</b>	<b>.16</b>	<b>.20</b>	<b>.11</b>	<b>.17</b>	<b>-.02</b>	<b>.16</b>	-	
39. Kge_36	.09	.10	.04	.04	-.05	-.07	-.07	.009	<b>.18</b>	<b>.34</b>	<b>.51</b>	<b>.59</b>	<b>.55</b>	<b>.68</b>	<b>.45</b>	<b>.34</b>	<b>.18</b>	<b>.36</b>	<b>.16</b>	-

**Anmerkungen.**

<sup>1</sup> Sechsstufige Antwortskala; die Nummerierung der Items entspricht der Nummerierung in Tabelle C.6.  
 Korrelationen innerhalb derselben Skala sind farblich markiert (Alter = rot; Körpergröße = grün;  
 Körpergewicht = blau); signifikante Ergebnisse ( $p < .01$ ) sind fett gedruckt.

Tabelle C.10.

*Spearman-Korrelationen ( $r_s$ ) Alter, Körpergröße, Körpergewicht und alle Items – Frauen, Bedingung 6PS<sup>1</sup>*

Merkmal und Item-Nummer	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1. Alter	-																		
2. Größe	-.11	-																	
3. Gewicht	.29	.37	-																
4. A_1	.24	.01	.18	-															
5. A_2	.35	-.02	.20	.37	-														
6. A_3	.60	-.08	.17	.24	.22	-													
7. A_4	.34	-.02	.17	.38	.61	.19	-												
8. A_5	.44	-.09	.20	.37	.65	.31	.59	-											
9. A_6	.09	-.008	.04	.16	.14	.12	.19	.19	-										
10. A_7	.69	-.13	.19	.30	.38	.47	.40	.47	.19	-									
11. A_8	.37	-.03	.09	.28	.39	.25	.41	.39	.22	.44	-								
12. A_9	.80	-.07	.26	.23	.31	.59	.25	.35	.10	.58	.31	-							
13. A_10	.61	-.03	.22	.32	.36	.36	.37	.38	.10	.63	.35	.48	-						
14. A_11	.29	.02	.14	.22	.41	.20	.34	.34	.21	.30	.28	.28	.23	-					
15. A_12	.43	-.05	.25	.39	.54	.27	.59	.58	.19	.50	.42	.33	.42	.37	-				
16. Kgr_13	-.10	.29	.14	.10	.10	-.02	.10	.03	.07	-.003	.05	-.08	-.03	.03	.11	-			
17. Kgr_14	-.03	.61	.22	.02	-.001	.02	.03	-.04	.01	-.06	-.006	-.03	-.06	.02	-.006	.32	-		
18. Kgr_15	-.03	.46	.21	.12	.11	-.009	.12	.08	.03	-.03	.07	-.05	.01	.04	.16	.30	.44	-	
19. Kgr_16	-.02	.75	.31	-.003	-.02	.006	-.03	-.05	-.009	-.05	-.02	.01	-.04	.02	-.03	.26	.62	.40	-
20. Kgr_17	.14	.46	.25	.09	.14	.10	.12	.10	-.05	.09	.11	.10	.11	.03	.12	.33	.45	.44	.43
21. Kgr_18	-.08	.46	.19	.04	.01	-.03	.06	-.01	.03	-.06	.07	-.11	.007	-.02	.09	.33	.47	.50	.38
22. Kgr_19	.04	.61	.31	.12	.07	.03	.07	.04	.05	.03	-.002	.02	.05	.06	.02	.33	.56	.46	.54
23. Kgr_20	-.13	.66	.22	.05	.05	-.08	.03	-.04	-.02	-.05	.07	-.07	-.03	.03	.05	.33	.55	.52	.58
24. Kgr_21	-.11	.71	.24	-.03	-.05	-.09	-.06	-.16	-.07	-.16	-.10	-.10	-.10	-.03	-.07	.24	.61	.39	.73
25. Kgr_22	.06	.58	.28	-.02	-.05	.09	-.07	-.07	-.06	-.01	-.07	.10	.000	-.006	-.08	.10	.46	.26	.58
26. Kgr_23	.05	.74	.38	.05	.03	.02	.02	-.04	.005	-.03	.03	.06	.03	.03	.007	.20	.53	.36	.69
27. Kgr_24	-.07	.64	.21	.04	.06	-.03	.04	-.02	.05	-.04	.06	-.05	-.02	-.01	.03	.34	.58	.50	.55
28. Kge_25	.04	.09	.25	.30	.26	.05	.24	.22	.12	.14	.21	.04	.14	.14	.26	.22	.10	.23	.03
29. Kge_26	-.14	.04	.29	.16	.16	-.10	.10	.11	.16	-.007	.06	-.06	-.009	.11	.15	.12	-.01	.08	.04
30. Kge_27	.11	.08	.57	.21	.22	.09	.13	.16	.007	.11	.12	.11	.14	.09	.27	.16	.11	.12	.08
31. Kge_28	.18	-.07	.58	.24	.27	.10	.21	.24	.07	.19	.13	.14	.18	.13	.29	.11	-.07	.03	-.07
32. Kge_29	.15	.01	.66	.24	.26	.09	.24	.27	.10	.18	.15	.13	.18	.17	.31	.15	-.003	.10	-.03
33. Kge_30	.16	-.04	.62	.25	.25	.13	.21	.25	.14	.19	.14	.15	.17	.16	.30	.14	-.06	-.007	-.06
34. Kge_31	.17	.06	.60	.14	.16	.13	.12	.17	-.008	.12	.09	.15	.15	.08	.22	.16	.08	.10	.03
35. Kge_32	.21	.06	.49	.19	.20	.10	.16	.21	.12	.19	.16	.15	.24	.08	.27	.06	-.03	.07	-.003
36. Kge_33	.15	.21	.40	.16	.04	.11	.05	.12	-.008	.14	.14	.12	.16	.06	.13	.10	.15	.15	.16
37. Kge_34	.10	.14	.47	.26	.26	.02	.22	.23	.12	.14	.07	.08	.15	.08	.22	.25	.16	.26	.09
38. Kge_35	-.03	-.03	.17	.10	.10	-.03	.10	.12	.10	.03	.06	-.06	-.004	.09	.12	.08	.02	.08	-.01
39. Kge_36	.03	-.05	.50	.15	.18	-.02	.20	.19	.07	.08	.07	.04	.09	.13	.19	.13	-.05	.03	-.03

Tabelle C.10. (Fortsetzung)

*Spearman-Korrelationen ( $r_s$ ) Alter, Körpergröße, Körpergewicht und alle Items – Frauen, Bedingung 6PS<sup>1</sup>*

Merkmal und Item-Nummer	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39
1. Alter																				
2. Größe																				
3. Gewicht																				
4. A_1																				
5. A_2																				
6. A_3																				
7. A_4																				
8. A_5																				
9. A_6																				
10. A_7																				
11. A_8																				
12. A_9																				
13. A_10																				
14. A_11																				
15. A_12																				
16. Kgr_13																				
17. Kgr_14																				
18. Kgr_15																				
19. Kgr_16																				
20. Kgr_17	-																			
21. Kgr_18	<b>.41</b>	-																		
22. Kgr_19	<b>.47</b>	<b>.51</b>	-																	
23. Kgr_20	<b>.41</b>	<b>.52</b>	<b>.59</b>	-																
24. Kgr_21	<b>.39</b>	<b>.37</b>	<b>.52</b>	<b>.53</b>	-															
25. Kgr_22	<b>.33</b>	<b>.26</b>	<b>.41</b>	<b>.37</b>	<b>.58</b>	-														
26. Kgr_23	<b>.41</b>	<b>.37</b>	<b>.53</b>	<b>.54</b>	<b>.65</b>	<b>.65</b>	-													
27. Kgr_24	<b>.43</b>	<b>.52</b>	<b>.55</b>	<b>.65</b>	<b>.52</b>	<b>.32</b>	<b>.49</b>	-												
28. Kge_25	<b>.19</b>	<b>.20</b>	<b>.17</b>	<b>.15</b>	.03	-.02	.05	<b>.16</b>	-											
29. Kge_26	.02	.10	<b>.13</b>	<b>.12</b>	-.02	-.06	.05	.10	<b>.25</b>	-										
30. Kge_27	<b>.13</b>	<b>.13</b>	<b>.12</b>	<b>.12</b>	.08	.02	.11	.09	<b>.30</b>	<b>.36</b>	-									
31. Kge_28	.009	.007	.06	.01	-.08	-.08	.01	-.07	<b>.34</b>	<b>.41</b>	<b>.65</b>	-								
32. Kge_29	.08	.08	.11	.02	-.05	-.05	.04	-.01	<b>.36</b>	<b>.39</b>	<b>.64</b>	<b>.78</b>	-							
33. Kge_30	.06	.008	.05	-.02	-.10	-.11	-.004	-.03	<b>.30</b>	<b>.44</b>	<b>.57</b>	<b>.75</b>	<b>.75</b>	-						
34. Kge_31	<b>.13</b>	<b>.12</b>	<b>.12</b>	.06	.10	<b>.13</b>	<b>.15</b>	.05	<b>.25</b>	<b>.26</b>	<b>.63</b>	<b>.57</b>	<b>.60</b>	<b>.55</b>	-					
35. Kge_32	.07	.02	.08	-.005	.03	.06	<b>.13</b>	.01	<b>.29</b>	<b>.26</b>	<b>.44</b>	<b>.46</b>	<b>.48</b>	<b>.44</b>	<b>.44</b>	-				
36. Kge_33	<b>.17</b>	<b>.16</b>	<b>.26</b>	<b>.16</b>	<b>.14</b>	<b>.22</b>	<b>.29</b>	<b>.12</b>	<b>.19</b>	<b>.12</b>	<b>.27</b>	<b>.23</b>	<b>.27</b>	<b>.26</b>	<b>.32</b>	<b>.31</b>	-			
37. Kge_34	<b>.22</b>	<b>.15</b>	<b>.28</b>	<b>.13</b>	<b>.12</b>	.03	<b>.15</b>	<b>.17</b>	<b>.41</b>	<b>.28</b>	<b>.51</b>	<b>.52</b>	<b>.58</b>	<b>.45</b>	<b>.43</b>	<b>.39</b>	<b>.29</b>	-		
38. Kge_35	-.03	.05	.04	.06	-.05	-.04	-.02	.06	<b>.13</b>	<b>.24</b>	<b>.16</b>	<b>.21</b>	<b>.20</b>	<b>.22</b>	<b>.16</b>	<b>.14</b>	<b>.11</b>	<b>.18</b>	-	
39. Kge_36	.08	-.007	.05	-.02	-.05	-.09	.01	-.04	<b>.22</b>	<b>.37</b>	<b>.45</b>	<b>.61</b>	<b>.62</b>	<b>.74</b>	<b>.42</b>	<b>.30</b>	<b>.25</b>	<b>.40</b>	<b>.13</b>	-

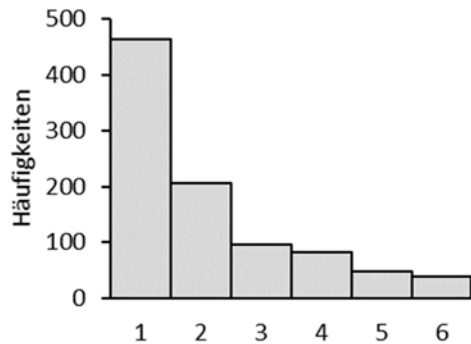
**Anmerkungen.**<sup>1</sup> Sechsstufige Antwortskala; die Nummerierung der Items entspricht der Nummerierung in Tabelle C.6.

Korrelationen innerhalb derselben Skala sind farblich markiert (Alter = rot; Körpergröße = grün;

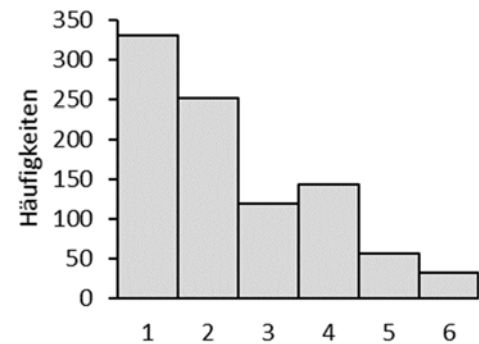
Körpergewicht = blau); signifikante Ergebnisse ( $p < .01$ ) sind fett gedruckt.

Abbildung C.14. Histogramme für das Merkmal Alter (Item 1-6) – 6PS

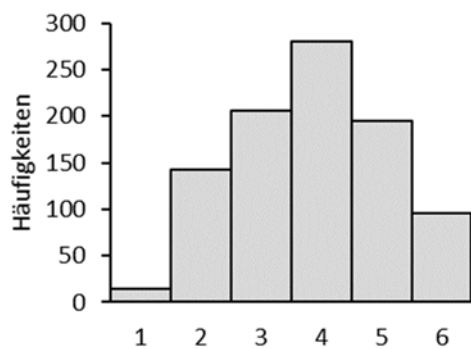
A\_1: Ich habe körperliche Gebrechen.



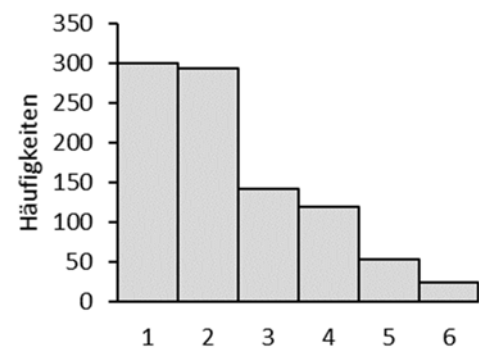
A\_2: Im Laufe der Zeit hat meine geistige Leistungsfähigkeit abgenommen.



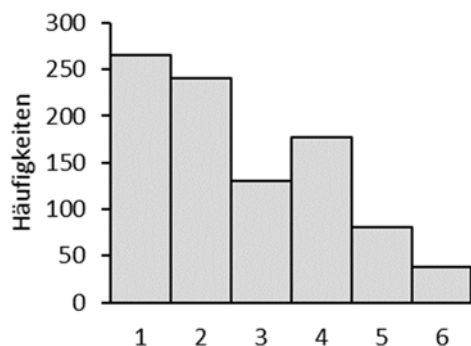
A\_3: Ich habe viel Lebenserfahrung.



A\_4: Im Laufe der Zeit hat meine Reaktionsfähigkeit abgenommen.



A\_5: Im Laufe der Zeit hat meine Merkfähigkeit abgenommen.



A\_6: Ich mache mir viele Gedanken über das Älterwerden.

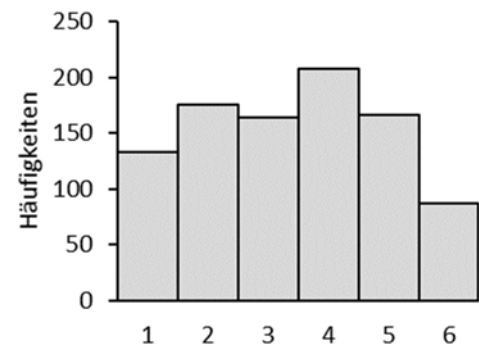
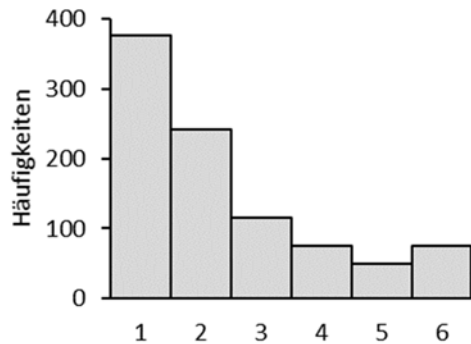


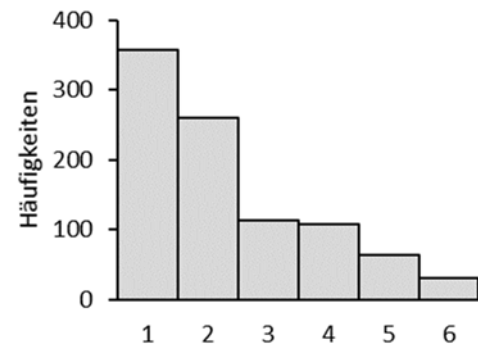


Abbildung C.15. Histogramme für das Merkmal Alter (Item 7-12) – 6PS

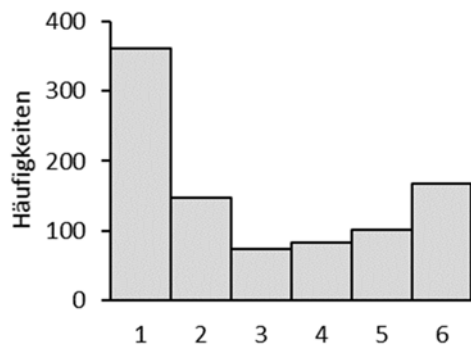
A\_7: Den größten Teil meines Lebens habe ich bereits gelebt.



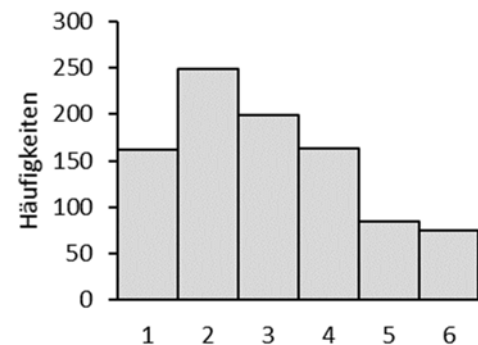
A\_8: Ich habe zunehmend Schwierigkeiten technischen Weiterentwicklungen zu folgen.



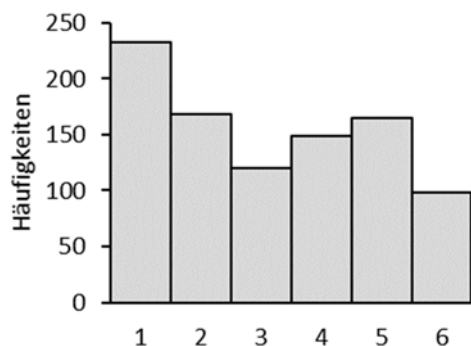
A\_9: Ich habe jahrelange Berufserfahrung.



A\_10: Mein ganzes Leben liegt noch vor mir.



A\_11: Früher war ich risikofreudiger.



A\_12: Ich verrichte Alltagstätigkeiten langsamer als früher.

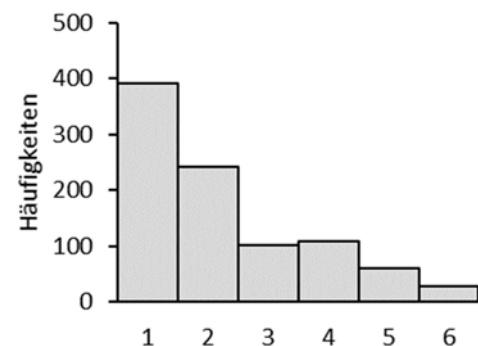
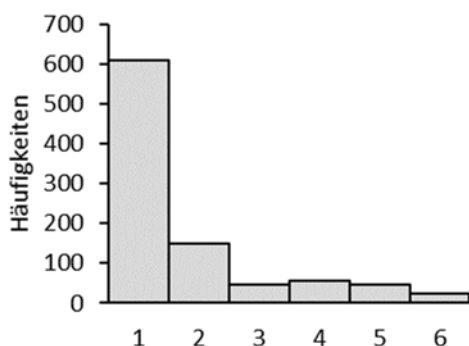
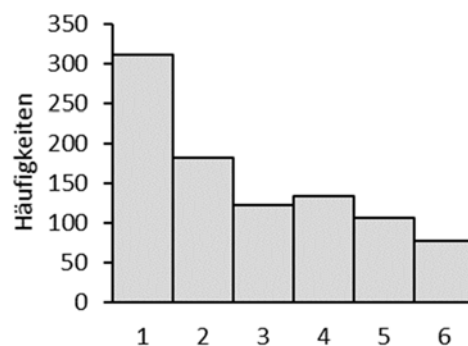


Abbildung C.16. Histogramme für das Merkmal Körpergröße (Item 13-18) – 6PS

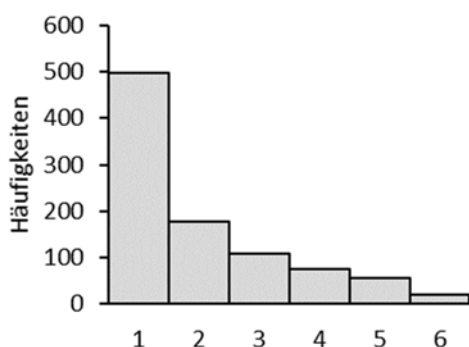
KGr\_13: Ich ziehe den Kopf ein, wenn ich durch Türen gehe.



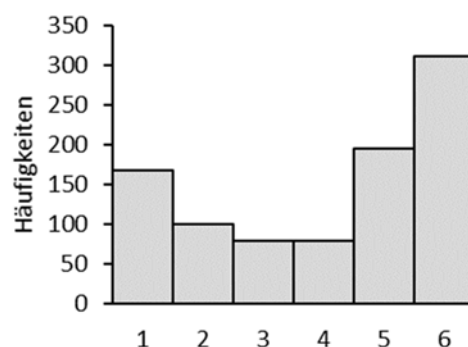
KGr\_14: Inmitten einer Menschenmenge kann ich über die meisten anderen Menschen hinwegblicken.



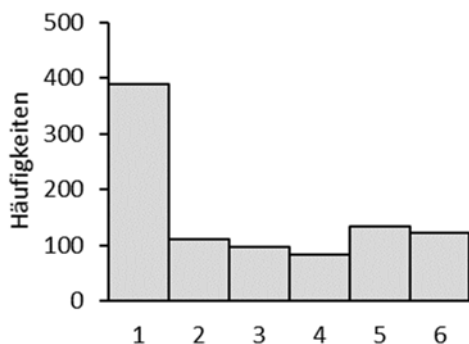
KGr\_15: Stühle und Tische sind meistens zu niedrig für mich.



KGr\_16: Im Supermarkt komme ich ohne Probleme an die Sachen im obersten Regal.



KGr\_17: Wenn ich ein Auto fahre, das vorher jemand anderes gefahren hat, muss ich üblicherweise den Sitz nach hinten schieben.



KGr\_18: Wenn ich mich in Hotelbetten gerade ausstrecke, dann stehen meine Füße über.

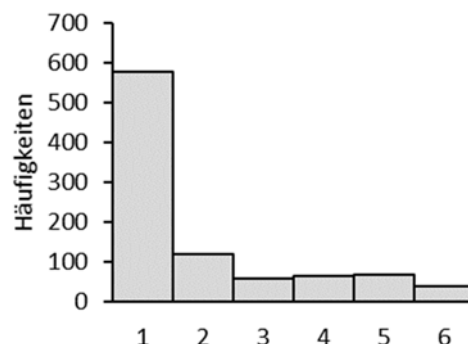
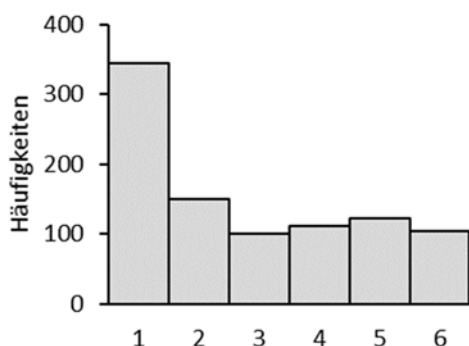
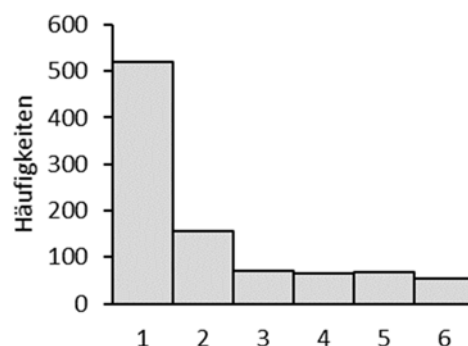


Abbildung C.17. Histogramme für das Merkmal Körpergröße (Item 19-24) – 6PS

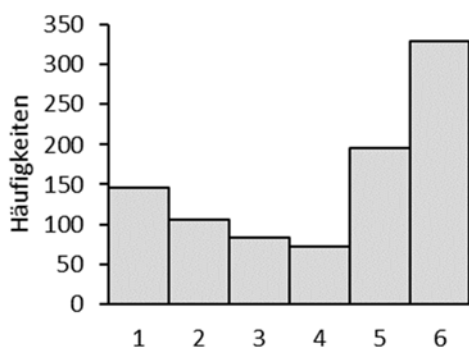
KGr\_19: Im Flugzeug stoße ich mit den Knien am Vordersitz an.



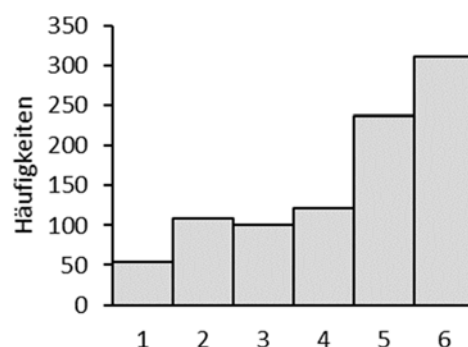
KGr\_20: Hosen sind mir häufig zu kurz.



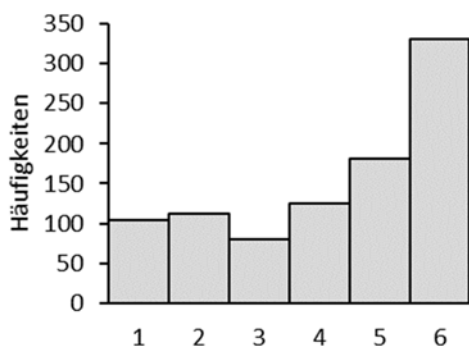
KGr\_21: Ich brauche einen Stuhl, wenn ich Sachen aus dem obersten Regal holen möchte.



KGr\_22: Wenn ich mich mit anderen Menschen unterhalte, muss ich nach oben schauen.



KGr\_23: Ich muss mich bei Gruppenfotos nach vorne stellen, damit ich gut zu sehen bin.



KGr\_24: Wenn ich andere Menschen zur Begrüßung umarme, muss ich mich nach unten beugen.

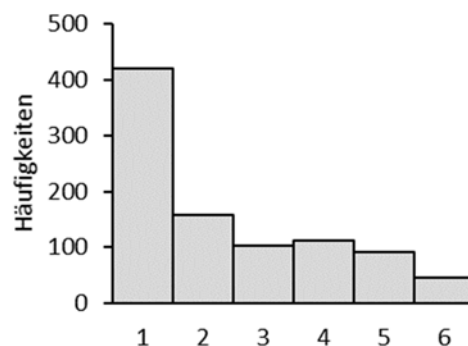
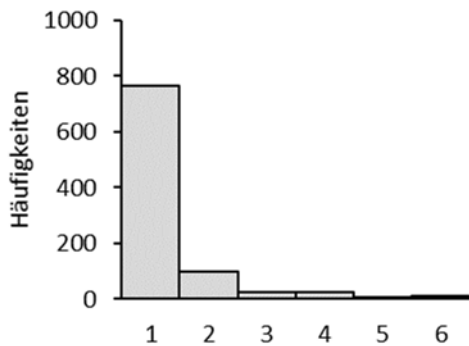
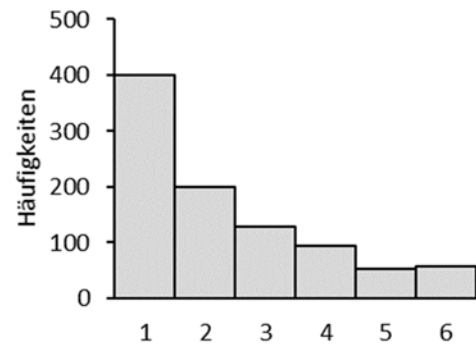


Abbildung C.18. Histogramme für das Merkmal Körpergewicht (Item 25-30) – 6PS

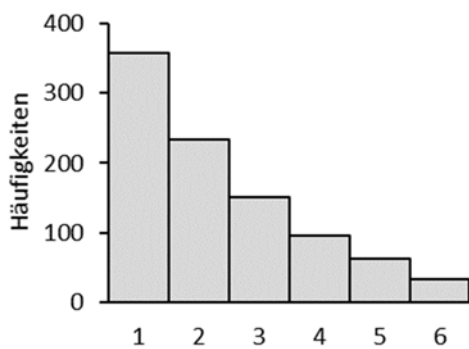
KGe\_25: Ich habe oft Angst, dass Stühle unter mir nachgeben.



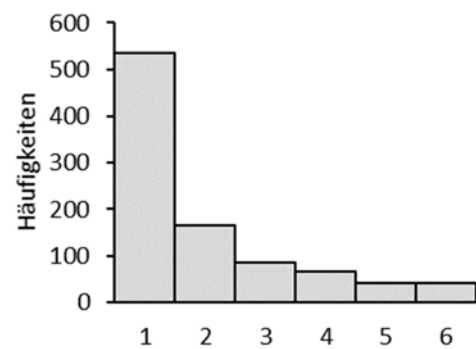
KGe\_26: Möglicherweise denken andere Menschen von mir, dass ich abends auf dem Sofa sitze und Schokolade und Chips esse.



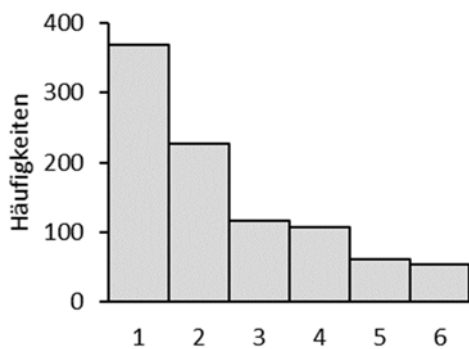
KGe\_27: Bei Hosen habe ich eine große Bundweite.



KGe\_28: Ich bin dick.



KGe\_29: Ich wiege viel.



KGe\_30: Ich müsste abnehmen.

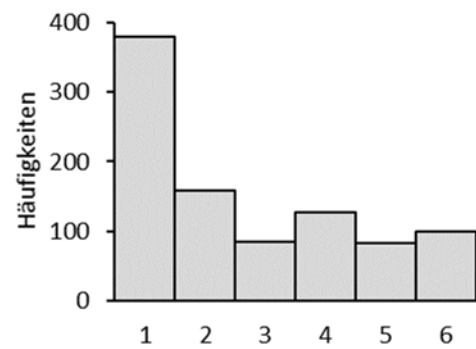
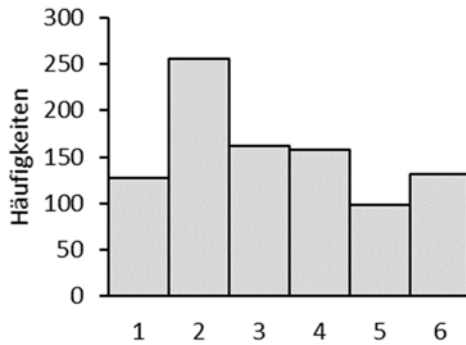
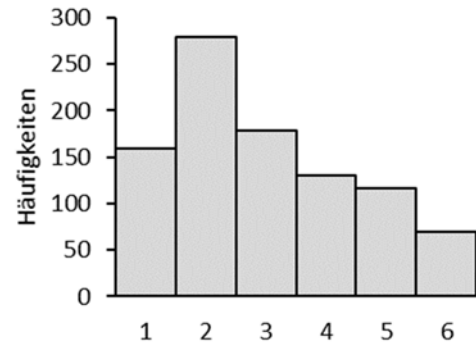


Abbildung C.19. Histogramme für das Merkmal Körpergewicht (Item 31-36) – 6PS

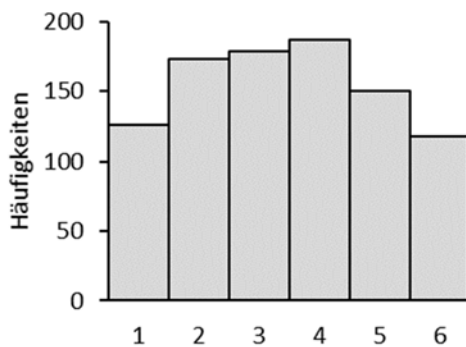
KGe\_31: Meine Hosen haben eine kleine Bundweite.



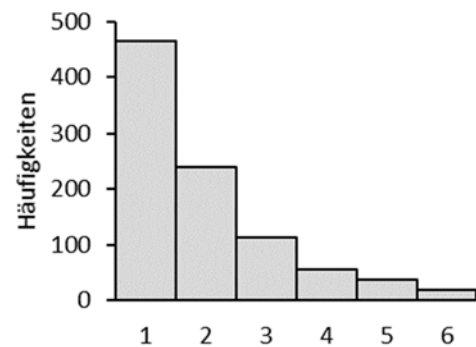
KGe\_32: Ich kann mich gut durch enge Spalten quetschen.



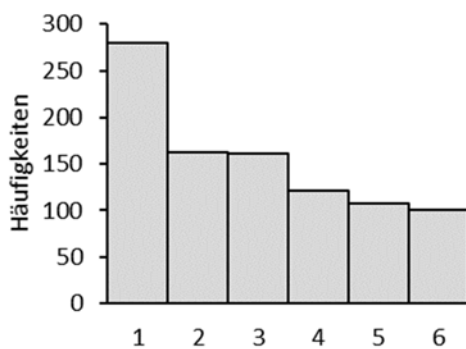
KGe\_33: Zwischen den Armlehnen in Flugzeugsitzen habe ich viel Platz.



KGe\_34: Im Aufzug nehme ich viel Platz ein.



KGe\_35: Wenn ich die Wahl habe zwischen Aufzug und Treppe, dann nehme ich den Aufzug.



KGe\_36: Ich sollte weniger essen.

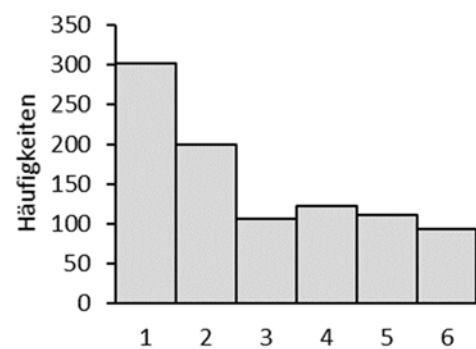


Tabelle C.11.

Explorative Faktorenanalyse mit sechs bzw. fünf Faktoren (Maximum Likelihood, Promax) – Bedingungen 2PS<sup>1</sup> und 6PS<sup>2</sup>

Merkmal und Item-Nummer	2PS <sup>1</sup>							6PS <sup>2</sup>					
	Faktor 1	Faktor 2	Faktor 3	Faktor 4	Faktor 5	Faktor 6	<i>h</i> <sup>2</sup>	Faktor 1	Faktor 2	Faktor 3	Faktor 4	Faktor 5	<i>h</i> <sup>2</sup>
Alter													
1 (D, P)		.25	<b>.32</b>				.268		.12	<b>.33</b>	.15	-.15	.327
2 (D)			<b>.97</b>	-.21			.693			<b>.92</b>		.14	.679
3 (D, P)			-.14	<b>.79</b>			.503				<b>.65</b>		.414
4 (D, P)	.11		<b>.77</b>	.13			.688			<b>.86</b>			.653
5 (D, P)		-.21	<b>1.07</b>	-.14			.866			<b>.88</b>			.665
6 (D, P)			<b>.18</b>			-.12	.050			<b>.26</b>			.080
7 (D)	-.13	-.15	.14	<b>.89</b>			.971			.23	<b>.76</b>	-.14	.795
8 (D, P)	-.13		.29	<b>.35</b>	-.15	.15	.422			<b>.46</b>	.22		.359
9 (D)		.12		<b>.81</b>		-.15	.641				<b>.75</b>		.657
10 (D, P)*			.17	<b>.69</b>			.694			.13	<b>.67</b>	-.11	.543
11 (D, P)			<b>.27</b>	.22			.196			<b>.50</b>			.274
12 (P)			<b>.66</b>	.13		.14	.577			<b>.69</b>	.12		.608
Körpergröße													
13 (D, P)	<b>.74</b>						.652	<b>.67</b>				-.22	.484
14 (D)	<b>.80</b>				.18		.717	<b>.85</b>				.15	.718
15 (D)	<b>.90</b>		.10				.731	<b>.80</b>	-.10		-.11	-.13	.655
16 (D, P)	.51				<b>.64</b>		.816	<b>.78</b>				.47	.757
17 (D, P)	<b>.74</b>			.18	.20		.664	<b>.75</b>			.13		.592
18 (D)	<b>.88</b>						.804	<b>.87</b>				-.17	.761
19 (D, P)	<b>.80</b>				.16		.760	<b>.83</b>					.693
20 (D, P)	<b>.75</b>	-.15					.500	<b>.74</b>					.532
21 (D, P)*	.51			-.12	<b>.65</b>		.838	<b>.73</b>			-.15	.53	.772
22 (D, P)*	.43			.15	<b>.67</b>		.787	.57				<b>.60</b>	.651
23 (D, P)*	.46				<b>.67</b>		.847	<b>.69</b>				.56	.746
24 (P)	<b>.88</b>			-.10	.14	-.11	.836	<b>.85</b>	-.12				.688
Körpergewicht													
25 (D)	.38	<b>.46</b>		.14	-.22		.457	.23	.33			<b>-.44</b>	.523
26 (D)		<b>.62</b>		-.32	-.15		.425		<b>.53</b>		-.18	-.14	.333
27 (D, P)		.47				<b>.75</b>	.995		<b>.80</b>				.670
28 (D)		<b>.89</b>		.15			.863	-.17	<b>.96</b>				.907
29 (D)		<b>.87</b>		.13		.14	.887		<b>.90</b>				.826
30 (D, P)	-.27	<b>.96</b>			.14		.866	-.25	<b>.91</b>				.807
31 (D, P)*	-.24	.47	.10	-.21	.39	<b>.61</b>	.828		<b>.80</b>	-.11	.11	.15	.576

Tabelle C.11. (Fortsetzung)

Explorative Faktorenanalyse mit sechs bzw. fünf Faktoren (Maximum Likelihood, Promax) – Bedingungen 2PS<sup>1</sup> und 6PS<sup>2</sup>

Merkmal und Item-Nummer	2PS <sup>1</sup>						$h^2$	6PS <sup>2</sup>					$h^2$
	Faktor 1	Faktor 2	Faktor 3	Faktor 4	Faktor 5	Faktor 6		Faktor 1	Faktor 2	Faktor 3	Faktor 4	Faktor 5	
32 (D)*		<b>.40</b>			.14	.31	.423		<b>.59</b>				.374
33 (D)*	<b>.27</b>	.19				.23	.261	.29	<b>.32</b>		.13		.235
34 (D)	<b>.36</b>	.33			-.19	.31	.540	.33	<b>.54</b>			-.12	.566
35 (P)		<b>.21</b>					.061		<b>.22</b>	.13	-.15		.094
36 (P)	-.22	<b>.84</b>				-.12	.604	-.22	<b>.79</b>				.576

Anmerkungen.

$h^2$  = Kommunalität des Items. Ladungen mit Betrag < .10 wurden unterdrückt.

Die stärkste Ladung für jedes Item ist fett gedruckt.

<sup>1</sup> Dichotome Antwortskala.

<sup>2</sup> Sechsstufige Antwortskala.

Die Nummerierung der Items entspricht der Nummerierung in Tabelle C.6.

D = Items hervorgegangen aus dem deduktiven Testkonstruktionsansatz.

P = Items hervorgegangen aus dem Prototypenansatz.

\* Items mit negativer Polung.

Tabelle C.12.

Faktorkorrelationen EFA mit sechs bzw. fünf Faktoren (Maximum Likelihood, Promax) – Bedingungen 2PS<sup>1</sup> und 6PS<sup>2</sup>

Faktor	2PS <sup>1</sup>						6PS <sup>2</sup>				
	1	2	3	4	5	6	1	2	3	4	5
1	-						-				
2	.23	-					.29	-			
3	-.02	.38	-				.10	.41	-		
4	-.06	.28	.65	-			.03	.18	.45	-	
5	.23	.01	.00	.11	-		-.05	-.32	-.45	.05	-
6	.31	.35	.10	.16	.10	-	-	-	-	-	-

Anmerkungen.

EFA = Explorative Faktorenanalyse.

<sup>1</sup> Dichotome Antwortskala.

<sup>2</sup> Sechsstufige Antwortskala.

Tabelle C.13.

*Unstandardisierte und standardisierte Ladungen CFA Modell 1 für die Bedingungen 2PS<sup>1</sup> und 6PS<sup>2</sup>*

Merkmal und Item-Nummer	2PS <sup>1</sup>			6PS <sup>2</sup>		
	$\lambda$ unstd.	$z$	$\lambda$ std.	$\lambda$ unstd.	$z$	$\lambda$ std.
Alter						
1 (D, P)	0.127	8.14	.341	0.764	13.22	.532
2 (D)	0.205	13.22	.485	0.961	21.07	.675
3 (D, P)	0.228	16.24	.459	0.517	11.09	.413
4 (D, P)	0.248	16.42	.589	0.927	21.12	.693
5 (D, P)	0.281	20.58	.595	1.032	22.58	.698
6 (D, P)	0.063	3.47	.128	0.390	6.94	.252
7 (D)	0.283	20.36	.689	1.128	21.50	.716
8 (D, P)	0.201	12.78	.483	0.758	15.02	.534
9 (D)	0.291	21.33	.592	1.097	18.60	.560
10 (D, P)*	0.308	24.17	.649	0.871	17.56	.583
11 (D, P)	0.180	10.92	.363	0.841	15.81	.488
12 (P)	0.219	14.44	.546	1.057	23.30	.743
Körpergröße						
13 (D, P)	0.116	9.56	.400	0.638	12.53	.480
14 (D)	0.315	25.78	.667	1.331	33.41	.792
15 (D)	0.149	11.35	.446	0.856	19.60	.627
16 (D, P)	0.305	26.01	.640	1.433	34.20	.747
17 (D, P)	0.324	26.98	.666	1.322	27.77	.698
18 (D)	0.209	15.53	.553	1.005	20.68	.658
19 (D, P)	0.331	29.46	.699	1.406	34.95	.779
20 (D, P)	0.161	11.69	.413	0.936	20.25	.597
21 (D, P)*	0.309	28.33	.648	1.360	32.47	.721
22 (D, P)*	0.220	17.66	.510	0.970	21.24	.611
23 (D, P)*	0.277	24.53	.602	1.267	29.58	.717
24 (P)	0.285	22.44	.655	1.195	28.26	.746
Körpergewicht						
25 (D)	0.047	4.27	.264	0.396	8.21	.444
26 (D)	0.118	6.61	.280	0.660	10.67	.430
27 (D, P)	0.237	15.27	.650	1.127	25.59	.792
28 (D)	0.187	11.82	.603	1.182	23.94	.827
29 (D)	0.298	21.57	.755	1.344	33.87	.880
30 (D, P)	0.311	24.28	.673	1.400	33.25	.791
31 (D, P)*	0.295	23.22	.593	1.110	25.13	.686
32 (D)*	0.255	16.46	.554	0.922	19.14	.608
33 (D)*	0.176	10.67	.351	0.665	13.08	.420



Tabelle C.13. (Fortsetzung)

*Unstandardisierte und standardisierte Ladungen CFA Modell 1 für die Bedingungen 2PS<sup>1</sup> und 6PS<sup>2</sup>*

Merkmal und Item-Nummer	2PS <sup>1</sup>			6PS <sup>2</sup>		
	$\lambda$ unstd.	$z$	$\lambda$ std.	$\lambda$ unstd.	$z$	$\lambda$ std.
34 (D)	0.100	7.07	.382	0.812	16.87	.651
35 (P)	0.084	4.64	.181	0.433	6.88	.253
36 (P)	0.234	15.17	.485	1.079	21.49	.624

*Anmerkungen.*

$\lambda$  unstd. = unstandardisierte Ladungen;  $\lambda$  std. = standardisierte Ladungen.

<sup>1</sup> Dichotome Antwortskala.

<sup>2</sup> Sechsstufige Antwortskala.

D = Items hervorgegangen aus dem deduktiven Testkonstruktionsansatz.

P = Items hervorgegangen aus dem Prototypenansatz.

\* Items mit negativer Polung; die Nummerierung der Items entspricht der Nummerierung in Tabelle C.6.

Alle Ladungen  $p \leq .001$ .

Abbildung C.20. CFA Modell 1 – 2PS

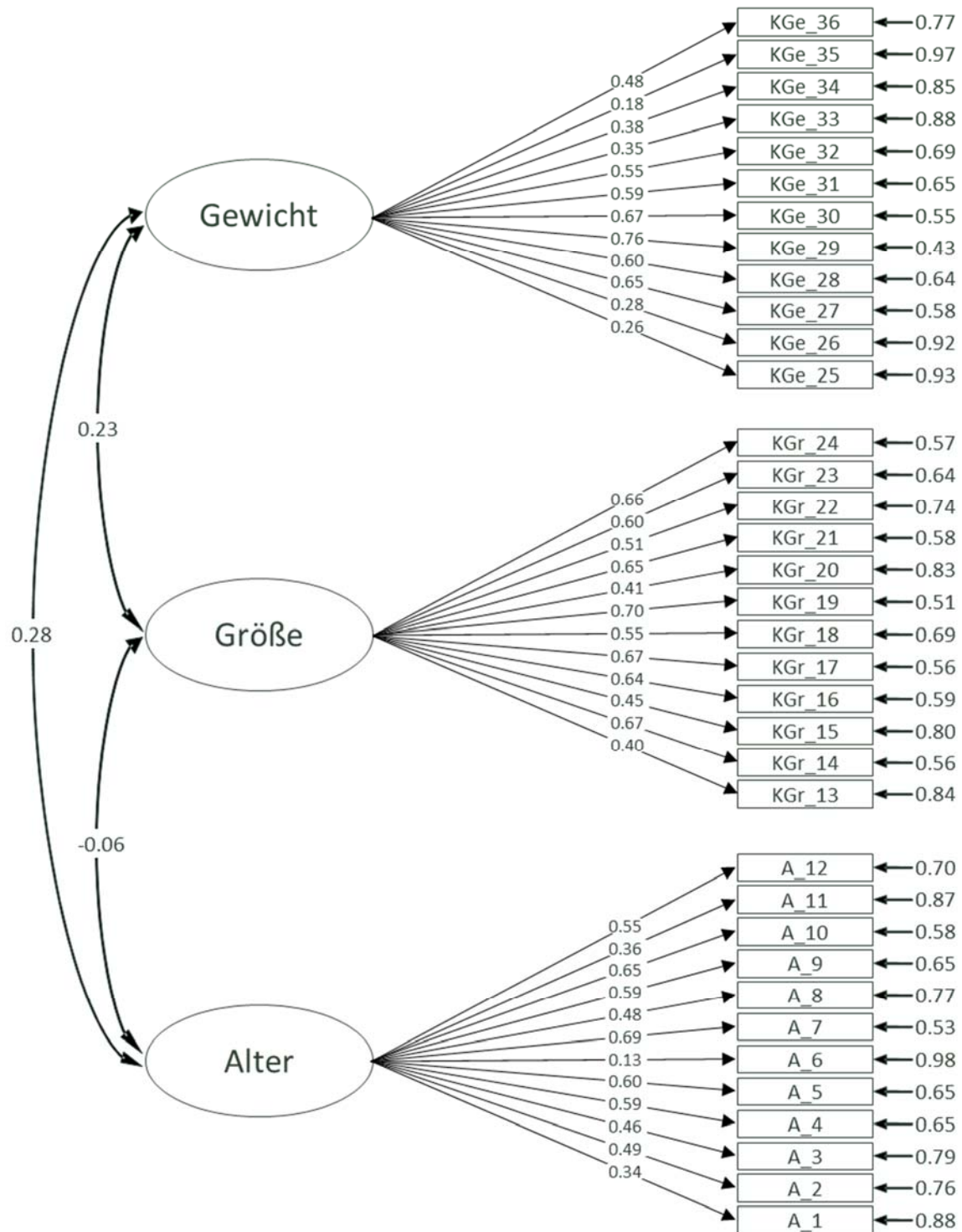


Abbildung C.21. CFA Modell 1 – 6PS

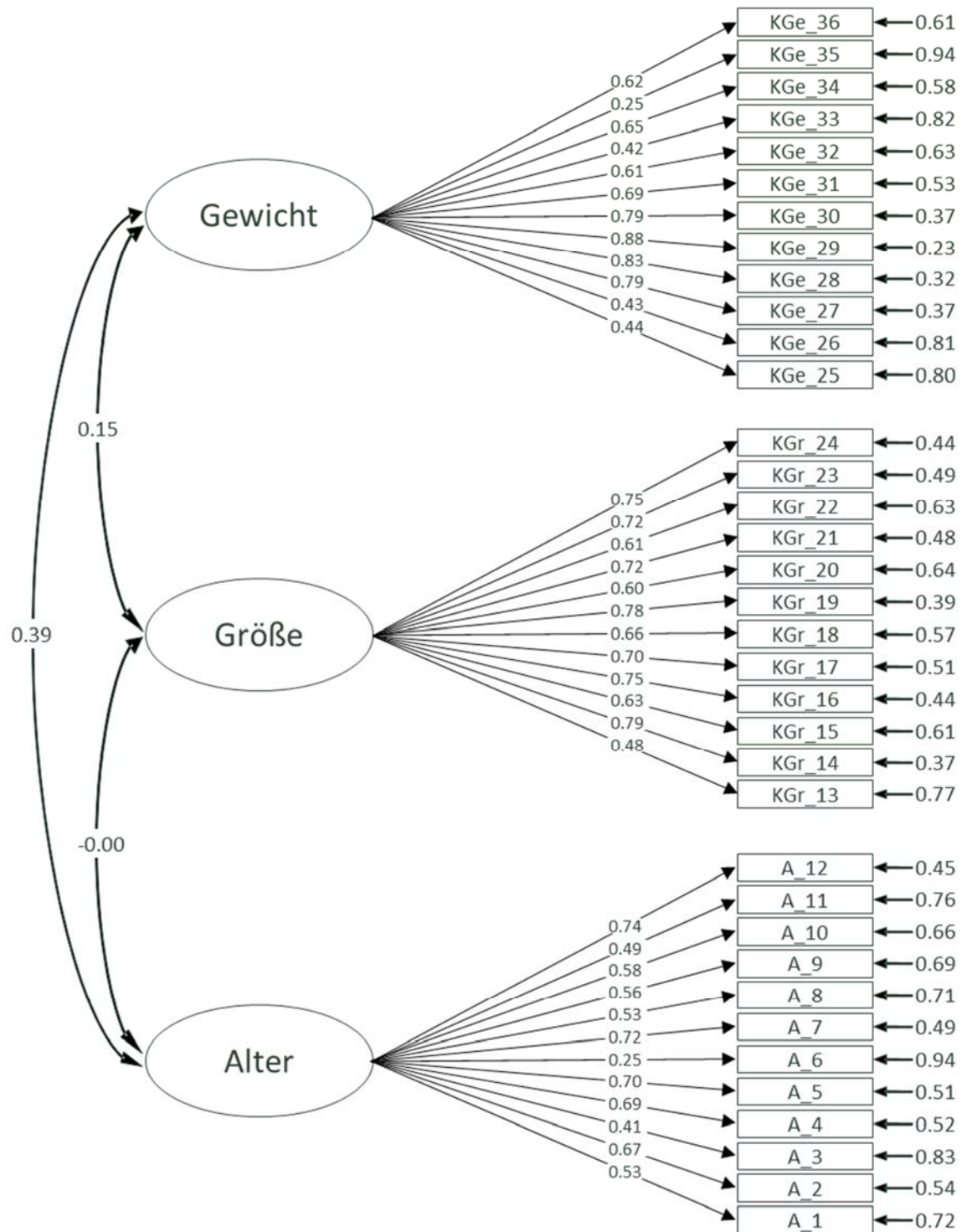


Tabelle C.14.

*Ergebnisse 1PL- und 2PL-Modell – Bedingung 2PS<sup>1</sup>*

Merkmal und Item- Nummer	1PL-Modell			2PL-Modell			
	$\sigma_i$	SE	z	$\sigma_i$	SE	z	$\beta_i$
Alter							
1 (D, P)	1.572	0.092	17.01	2.116	0.25	8.46	0.877
2 (D)	1.163	0.081	14.44	1.119	0.097	11.53	1.454
3 (D, P)	-0.223	0.066	-3.38	-0.227	0.07	-3.23	1.256
4 (D, P)	1.186	0.081	14.62	0.965	0.073	13.31	2.065
5 (D, P)	0.680	0.071	9.60	0.572	0.062	9.21	1.934
6 (D, P)	0.294	0.067	4.42	1.300	0.506	2.57	0.237
7 (D)	1.265	0.083	15.20	0.891	0.059	15.01	3.287
8 (D, P)	1.228	0.082	14.94	1.247	0.110	11.34	1.323
9 (D)	0.352	0.067	5.25	0.306	0.059	5.23	1.827
10 (D, P)*	0.636	0.070	9.07	0.504	0.057	8.89	2.265
11 (D, P)	0.263	0.066	3.97	0.359	0.093	3.87	0.887
12 (P)	1.346	0.085	15.76	1.175	0.089	13.23	1.751
Körpergröße							
13 (D, P)	1.563	0.075	20.76	1.679	0.109	15.46	2.227
14 (D)	0.500	0.050	10.04	0.491	0.053	9.33	2.538
15 (D)	1.343	0.068	19.86	1.410	0.088	15.95	2.329
16 (D, P)	-0.459	0.050	-9.15	-0.428	0.047	-9.12	3.410
17 (D, P)	0.340	0.048	7.06	0.351	0.053	6.65	2.156
18 (D)	1.120	0.061	18.24	1.071	0.065	16.44	3.089
19 (D, P)	0.486	0.050	9.79	0.465	0.051	9.19	2.725
20 (D, P)	1.051	0.060	17.60	1.386	0.114	12.18	1.410
21 (D, P)*	-0.455	0.050	-9.08	-0.422	0.047	-8.97	3.514
22 (D, P)*	-0.820	0.056	-14.71	-0.794	0.057	-13.84	2.836
23 (D, P)*	-0.612	0.052	-11.71	-0.572	0.049	-11.64	3.329
24 (P)	0.784	0.054	14.51	0.728	0.053	13.62	3.230
Körpergewicht							
25 (D)	2.989	0.166	17.95	3.050	0.373	8.18	1.400
26 (D)	1.130	0.078	14.51	1.573	0.175	9.00	0.900
27 (D, P)	1.551	0.090	17.18	1.244	0.074	16.81	2.485
28 (D)	1.944	0.105	18.47	1.351	0.057	23.52	4.755
29 (D)	1.333	0.083	15.98	0.924	0.058	16.05	5.279
30 (D, P)	0.751	0.070	10.77	0.580	0.051	11.44	3.332
31 (D, P)*	0.159	0.063	2.52	0.168	0.057	2.93	1.835
32 (D)*	0.765	0.070	10.94	0.843	0.086	9.76	1.303

Tabelle C.14. (Fortsetzung)

Ergebnisse 1PL- und 2PL-Modell – Bedingung 2PS<sup>1</sup>

Merkmal und Item- Nummer	1PL-Modell			2PL-Modell			
	$\sigma_i$	$SE$	$z$	$\sigma_i$	$SE$	$z$	$\beta_i$
33 (D)*	0.026	0.063	0.42	0.083	0.113	0.73	0.642
34 (D)	2.295	0.122	18.84	2.407	0.234	10.31	1.344
35 (P)	0.746	0.070	10.72	2.116	0.468	4.52	0.393
36 (P)	0.511	0.066	7.73	0.504	0.065	7.79	1.670

Anmerkungen.

$\sigma_i$  = Schwierigkeitsparameter;  $SE$  = Standardfehler;  $\beta_i$  = Diskriminationsparameter.

$\widehat{SD}(\theta)$  für das 1PL-Modell = 1.358 (Alter), 2.634 (Körpergröße) und 1.453 (Körpergewicht).

<sup>1</sup>Dichotome Antwortskala; die Nummerierung der Items entspricht der Nummerierung in Tabelle C.6.

D = Items hervorgegangen aus dem deduktiven Testkonstruktionsansatz.

P = Items hervorgegangen aus dem Prototypenansatz.

\* Items mit negativer Polung.

Tabelle C.15.

*Ergebnisse Partial-Credit-Modell – Bedingung 6PS<sup>1</sup>*

Merkmal und Item- Nummer	t <sub>1</sub>	SE <sub>t1</sub>	t <sub>2</sub>	SE <sub>t2</sub>	t <sub>3</sub>	SE <sub>t3</sub>	t <sub>4</sub>	SE <sub>t4</sub>	t <sub>5</sub>	SE <sub>t5</sub>
<b>Alter</b>										
1 (D, P)	1.227	0.162	1.551	0.230	0.775	0.273	1.865	0.331	1.661	0.396
2 (D)	0.148	0.157	1.365	0.207	0.095	0.225	2.474	0.293	2.167	0.405
3 (D, P)	-4.979	0.508	-1.208	0.203	-0.687	0.171	0.908	0.174	1.961	0.235
4 (D, P)	-0.296	0.154	1.355	0.192	0.747	0.227	2.252	0.303	2.666	0.449
5 (D, P)	-0.264	0.166	1.036	0.201	-0.222	0.211	2.077	0.251	2.414	0.362
6 (D, P)	-1.185	0.213	-0.206	0.199	-0.386	0.192	0.804	0.192	1.935	0.247
7 (D)	0.481	0.155	1.383	0.210	1.224	0.271	1.460	0.332	0.378	0.332
8 (D, P)	0.262	0.153	1.582	0.211	0.557	0.245	1.770	0.291	2.463	0.402
9 (D)	1.154	0.186	1.134	0.262	-0.036	0.290	0.129	0.268	-0.079	0.231
10 (D, P)*	-1.326	0.189	0.226	0.176	0.551	0.194	1.722	0.248	1.159	0.290
11 (D, P)	0.015	0.187	0.383	0.218	-0.267	0.223	0.275	0.207	1.743	0.237
12 (P)	0.579	0.155	1.671	0.222	0.345	0.251	1.880	0.294	2.543	0.414
<b>Körpergröße</b>										
13 (D, P)	1.429	0.109	1.663	0.183	0.600	0.212	1.354	0.216	2.245	0.280
14 (D)	0.077	0.108	0.423	0.129	0.270	0.137	0.943	0.143	1.389	0.167
15 (D)	0.921	0.103	0.839	0.134	1.071	0.163	1.388	0.194	2.572	0.291
16 (D, P)	-0.439	0.143	-0.280	0.164	-0.142	0.172	-0.767	0.148	0.056	0.104
17 (D, P)	0.883	0.124	0.151	0.151	0.513	0.162	0.117	0.153	1.045	0.140
18 (D)	1.530	0.119	1.119	0.173	0.542	0.193	0.945	0.188	1.902	0.221
19 (D, P)	0.404	0.112	0.411	0.141	0.239	0.149	0.537	0.143	1.167	0.149
20 (D, P)	1.092	0.107	1.146	0.157	0.709	0.185	0.908	0.189	1.541	0.205
21 (D, P)*	-0.678	0.145	-0.343	0.161	-0.037	0.174	-0.905	0.154	-0.030	0.103
22 (D, P)*	-2.070	0.192	-0.765	0.155	-0.563	0.149	-0.655	0.126	0.190	0.099
23 (D, P)*	-1.223	0.156	-0.321	0.161	-0.733	0.158	-0.276	0.129	-0.131	0.106
24 (P)	0.715	0.108	0.610	0.139	0.415	0.148	1.057	0.154	1.953	0.203
<b>Körpergewicht</b>										
25 (D)	3.285	0.200	2.875	0.373	1.028	0.461	3.486	0.681	1.052	0.745
26 (D)	0.759	0.145	0.777	0.183	0.913	0.217	1.773	0.277	1.111	0.308
27 (D, P)	0.332	0.139	0.745	0.170	1.170	0.211	1.534	0.261	2.341	0.348
28 (D)	1.655	0.155	1.260	0.215	0.986	0.260	1.814	0.320	1.390	0.357
29 (D)	0.425	0.140	1.088	0.186	0.535	0.214	1.703	0.258	1.459	0.303
30 (D, P)	0.952	0.158	0.946	0.216	-0.370	0.224	1.310	0.227	0.733	0.239
31 (D, P)*	-1.753	0.181	0.431	0.163	0.115	0.179	1.189	0.208	0.392	0.215
32 (D)*	-1.452	0.166	0.515	0.157	0.680	0.186	0.745	0.205	1.820	0.247
33 (D)*	-1.227	0.191	-0.412	0.172	-0.068	0.169	0.720	0.177	1.178	0.201
34 (D)	0.850	0.135	1.395	0.187	1.749	0.262	1.743	0.335	2.662	0.453

Tabelle C.15. (Fortsetzung)

*Ergebnisse Partial-Credit-Modell – Bedingung 6PS<sup>1</sup>*

Merkmal und Item- Nummer	t <sub>1</sub>	SE <sub>t1</sub>	t <sub>2</sub>	SE <sub>t2</sub>	t <sub>3</sub>	SE <sub>t3</sub>	t <sub>4</sub>	SE <sub>t4</sub>	t <sub>5</sub>	SE <sub>t5</sub>
35 (P)	0.337	0.162	-0.128	0.178	0.643	0.193	0.748	0.213	1.051	0.225
36 (P)	0.177	0.150	0.872	0.194	0.013	0.212	0.735	0.210	1.248	0.228

*Anmerkungen.*

t<sub>1</sub> = erste Schwelle; t<sub>2</sub> = zweite Schwelle; t<sub>3</sub> = dritte Schwelle; t<sub>4</sub> = vierte Schwelle; t<sub>5</sub> = fünfte Schwelle;

SE = Standardfehler.

$\widehat{SD}(\theta)$  für das Partial-Credit-Modell = 0.561 (Alter), 0.950 (Körpergröße) und 0.642 (Körpergewicht).

<sup>1</sup> Sechsstufige Antwortskala; die Nummerierung der Items entspricht der Nummerierung in Tabelle C.6.

D = Items hervorgegangen aus dem deduktiven Testkonstruktionsansatz.

P = Items hervorgegangen aus dem Prototypenansatz.

\* Items mit negativer Polung.

Tabelle C.16.

*Ergebnisse Generalisiertes-Partial-Credit Modell – Bedingung 6PS<sup>1</sup>*

Merkmal und Item-Nummer	t <sub>1</sub>	SE <sub>t1</sub>	t <sub>2</sub>	SE <sub>t2</sub>	t <sub>3</sub>	SE <sub>t3</sub>	t <sub>4</sub>	SE <sub>t4</sub>	t <sub>5</sub>	SE <sub>t5</sub>	β <sub>i</sub>
<b>Alter</b>											
1 (D, P)	1.727	0.276	1.976	0.336	0.779	0.356	2.037	0.439	1.586	0.516	0.425
2 (D)	-0.142	0.095	0.842	0.123	0.393	0.129	1.863	0.172	1.922	0.228	1.046
3 (D, P)	-6.401	0.851	-1.369	0.284	-0.861	0.243	1.129	0.249	2.333	0.349	0.402
4 (D, P)	-0.396	0.089	0.824	0.112	0.765	0.123	1.759	0.169	2.218	0.246	1.084
5 (D, P)	-0.453	0.094	0.560	0.114	0.172	0.117	1.561	0.141	1.983	0.197	1.100
6 (D, P)	-1.927	0.716	0.250	0.651	-1.385	0.659	1.520	0.645	4.167	0.955	0.169
7 (D)	0.115	0.104	0.924	0.135	1.019	0.162	1.298	0.196	0.831	0.203	0.978
8 (D, P)	0.264	0.159	1.583	0.233	0.573	0.241	1.739	0.289	2.392	0.399	0.571
9 (D)	1.852	0.326	1.661	0.391	-0.128	0.403	-0.066	0.376	-0.544	0.344	0.402
10 (D, P)*	-1.306	0.184	0.228	0.171	0.568	0.185	1.683	0.244	1.138	0.276	0.591
11 (D, P)	0.412	0.291	0.737	0.333	-0.458	0.335	0.116	0.308	2.042	0.358	0.375
12 (P)	0.127	0.094	1.031	0.129	0.576	0.136	1.557	0.161	2.122	0.226	1.091
<b>Körpergröße</b>											
13 (D, P)	2.417	0.276	2.380	0.342	0.318	0.357	1.334	0.355	2.543	0.459	0.563
14 (D)	-0.122	0.083	0.311	0.092	0.328	0.097	0.906	0.101	1.395	0.122	1.394
15 (D)	0.954	0.135	0.843	0.141	1.079	0.169	1.414	0.201	2.650	0.306	0.920
16 (D, P)	-0.572	0.122	-0.338	0.133	-0.158	0.138	-0.582	0.129	0.175	0.091	1.205
17 (D, P)	1.337	0.203	0.167	0.198	0.531	0.213	-0.090	0.209	1.029	0.183	0.714
18 (D)	1.309	0.148	1.033	0.154	0.587	0.169	0.993	0.165	1.911	0.196	1.102
19 (D, P)	0.311	0.116	0.368	0.129	0.249	0.135	0.557	0.130	1.188	0.137	1.058
20 (D, P)	1.500	0.193	1.367	0.220	0.675	0.244	0.847	0.249	1.568	0.267	0.715
21 (D, P)*	-0.712	0.137	-0.364	0.149	-0.059	0.160	-0.823	0.155	0.024	0.103	1.042
22 (D, P)*	-2.146	0.234	-0.689	0.188	-0.579	0.181	-0.819	0.164	0.097	0.124	0.770
23 (D, P)*	-1.226	0.143	-0.362	0.148	-0.693	0.145	-0.246	0.118	-0.056	0.105	1.062
24 (P)	0.468	0.102	0.529	0.109	0.468	0.115	1.052	0.122	1.893	0.165	1.244
<b>Körpergewicht</b>											
25 (D)	4.400	0.571	3.531	0.607	0.703	0.637	3.873	0.963	0.347	1.049	0.465
26 (D)	1.563	0.302	1.207	0.330	1.072	0.377	2.191	0.490	0.617	0.533	0.368
27 (D, P)	-0.138	0.082	0.417	0.093	0.864	0.114	1.282	0.145	2.024	0.199	1.332
28 (D)	0.226	0.061	0.666	0.062	0.854	0.075	1.331	0.105	1.680	0.125	3.208
29 (D)	-0.252	0.055	0.403	0.069	0.577	0.070	1.149	0.096	1.541	0.121	2.535
30 (D, P)	0.030	0.083	0.370	0.098	0.066	0.101	0.931	0.105	1.004	0.119	1.620
31 (D, P)*	-1.634	0.143	0.169	0.131	0.074	0.136	0.970	0.162	0.512	0.170	0.872
32 (D)*	-1.646	0.201	0.594	0.197	0.695	0.222	0.660	0.242	1.802	0.296	0.543
33 (D)*	-1.740	0.469	-0.416	0.417	-0.232	0.410	1.018	0.437	1.326	0.488	0.257
34 (D)	0.723	0.152	1.274	0.193	1.618	0.257	1.632	0.322	2.549	0.436	0.681



Tabelle C.16. (Fortsetzung)

*Ergebnisse Generalisiertes-Partial-Credit Modell – Bedingung 6PS<sup>1</sup>*

Merkmal und Item-Nummer	$t_1$	$SE_{t1}$	$t_2$	$SE_{t2}$	$t_3$	$SE_{t3}$	$t_4$	$SE_{t4}$	$t_5$	$SE_{t5}$	$\beta_i$
35 (P)	3.594	0.952	-0.039	0.793	2.041	0.915	1.018	0.954	0.691	0.990	0.140
36 (P)	-0.049	0.132	0.636	0.168	0.026	0.173	0.669	0.174	1.171	0.192	0.799

*Anmerkungen.*

$t_1$  = erste Schwelle;  $t_2$  = zweite Schwelle;  $t_3$  = dritte Schwelle;  $t_4$  = vierte Schwelle;  $t_5$  = fünfte Schwelle;

$SE$  = Standardfehler;  $\beta_i$  = Diskriminationsparameter.

<sup>1</sup> Sechsstufige Antwortskala; die Nummerierung der Items entspricht der Nummerierung in Tabelle C.6.

D = Items hervorgegangen aus dem deduktiven Testkonstruktionsansatz.

P = Items hervorgegangen aus dem Prototypenansatz.

\* Items mit negativer Polung.